# Integrative analysis of morphology, multi-locus genotyping and host usage – a case study in *Eimeria* spp., intracellular parasites of rodents

**D i s s e r t a t i o n**

zur Erlangung des akademischen Grades

Doctor in Philosophy (PhD)

im Fach Biologie

Eingereicht an der
Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin

von
***Víctor Hugo Jarquín-Díaz***

aus San Cristóbal de las Casas, Chiapas, Mexiko

Präsidentin der Humboldt-Universität zu Berlin
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Lebenswissenschaftlichen Fakultät
Prof. Dr. Bernhard Grimm

Gutachter: 1. Prof. Dr Emanuel Heitlinger
2. Prof. Dr. Alex Greenwood
3. Prof. Dr. Dino McMahon

Datum der Einreichung: 02.02.2021

*“Counting how many kinds there are can be difficult, because some places are hard to look in… but mainly it’s difficult because there are just so many of them”*

*Nicola Davies*

# Zusammenfassung

Die Identifizierung von Arten stellt eine der größten Herausforderungen in der Biologie dar. Methodische Ansätze zur Identifizierung erfüllen nicht immer die für jedes Konzept erforderlichen Merkmale. Bei komplexen Organismen, wie z.B. eukaryotischen Parasiten, gehen herkömmliche Ansätze kaum auf eine strenge Prüfung der morphologischen Konsistenz und Persistenz in verschiedenen Umgebungen ein. Die Einbeziehung der Wirtsspezifität könnte die Identifizierung in die Irre führen. Molekulare Ansätze stellen eine einfache Alternative für die Artbestimmung im Rahmen eines phlyogenetischen Artkonzeptes dar. Kokzidien werden hier als Beispiel für die morphologische Identifizierung - in vielen Fällen sogar Beschreibungen - verwendet, die durch Annahmen zur Wirtsspezifität verfälscht werden. Diese Dissertation kombiniert verschiedene Ansätze zur Parasitenbestimmung, die es erlauben, die Vielfalt der Parasiten in natürlichen Systemen zu beurteilen. Ich konzentriere mich insbesondere darauf, wie die Artbestimmung in der Gattung *Eimeria* mit der Wirtsspezifität bei Nagetierarten zusammenhängt. Zunächst bietet diese Arbeit eine Reihe von Methoden zur Beurteilung der Prävalenz auf der Ebene der Parasitenarten in *Mus musculus* (Hausmäuse). Der Ansatz integriert die morphologische Beschreibung mit molekularen Methoden zum Nachweis, zur Nischenannäherung und zur phylogenetischen Rekonstruktion. Als Ergebnis war es möglich, drei verschiedene *Eimeria*-Spezies zu identifizieren, Mäuse mit Doppelinfektionen zu erkennen und die artenspezifische Prävalenz in Abhängigkeit von der Wirtsdichte vorherzusagen. Zur Identifizierung von *Eimeria* spp. über verschiedene Wirtsarten hinweg wurde eine neuartige Hochdurchsatz-Multi-Locus-Genotypisierungsmethode etabliert und mit der auf zuvor etablierten Markern basierenden Einzelmarker-Genotypisierung verglichen. Der Multi-Locus-Genotypisierungsansatz lieferte eine höhere Auflösung und ermöglichte die Unterscheidung eng verwandter *Eimeria*-Isolate. Dies bestätigte, dass die Art *E. falciformis* in einer einzigen Wirtsart, der Hausmaus, vorkommt. *E. vermiformis* und *E. apionodes* konnten jedoch nicht unterschieden werden, was auf eine einzige Art mit breitem Wirtsspektrum hindeutet. *E. vermiformis* und E. *apionodes* konnten jedoch nicht unterschieden werden, was auf eine einzige Art mit breiter Wirtsverwendung in einem phylogenetischen Artkonzept schließen lässt. Diese Ergebnisse zeigen, dass die hohe Wirtsspezifität, die traditionell für *Eimeria*-Parasiten angenommen wird, fragwürdig ist und dass die Identifizierung von Arten durch Wirtsassoziation vermieden werden sollte.
Die hier entwickelten Ansätze zur Identifizierung von *Eimeria* spp. erlaubten die Differenzierung eng verwandter Isolate mit nicht unterscheidbarer Morphologie. Durch molekulare Amplifikation, Sequenzierung, Genotypisierung und phylogenetische Analyse war es möglich, Eimerien auf Artniveau zu identifizieren und die Wirtsspezifität in Isolaten aus natürlichen Systemen in Frage zu stellen. In einer breiteren Perspektive betonte diese Arbeit die Notwendigkeit, Strategien bei der Erkennung, Quantifizierung und Identifizierung von Parasiten zu standardisieren und zu kombinieren, um ein besseres Verständnis auf evolutionärer und ökologischer Ebene zu erlangen.

**Schlagwörter:** Eimerien, Nagetiere, Morphologie, Phylogenetik, Genotypisierung

# Abstract

The identification of species represents one of the major challenges in biology. However, methodological approaches for identification are not always fulfilling the characteristics stated for each concept. In complex organisms, such as eukaryotic parasites, conventional approaches hardly address a strict test to maintain morphological consistency and persistence between different environments. Moreover, incorporating host specificity might mislead the identification. Molecular approaches represent a straightforward alternative for species identification under a phylogenetic species concept. Coccidia exemplifies morphological identification confounded by assumptions of host specificity. This PhD thesis combines different approaches for parasite identification to assess the diversity of parasites in natural systems. Particularly, I focus on how species identification in the genus *Eimeria* is linked to its host specificity in rodent species. First, this thesis provides a set of methods to assess prevalence at the species level in *Mus musculus* (house mice) systems. The approach integrates morphological description with molecular methods for detection, niche approximation and phylogenetic reconstruction. As a result, three different *Eimeria* species were identified, mice with double infections were detected and species-specific prevalence were predicted to be host density-dependent. For identification of *Eimeria* spp. across different host species, a novel high-throughput multi-locus genotyping was established and compared with single-marker genotyping. The multi-locus genotyping approach provided a higher resolution to distinguish closely related *Eimeria* isolates. This confirmed the species *E. falciformis* to have a single host species, the house mice. However, *E. vermiformis* and *E. apionodes* could not be distinguished suggesting a single species with broader host usage in a phylogenetic species concept. These findings show that the high host specificity traditionally assumed for *Eimeria* parasites is questionable, and that identification of species by host association should be avoided.

The approaches for identification of *Eimeria* spp. Developed here allowed differentiation of closely related isolates with indistinguishable morphology. Molecular amplification, sequencing, genotyping and phylogeny allowed the identification of *Eimeria* at species level and to question host specificity in isolates from natural systems. In a broader perspective, this work emphasised the necessity to standardise and combine strategies in parasite detection, quantification and identification to gain better understanding at an evolutionary and ecological level.



**Keywords:**
*Eimeria*, Rodents, Morphology, Phylogenetics, Genotyping

# Table of Content

# List of Figures



# List of Tables

# List of Abbreviations

| Abbreviation | Full term |
|---|---|
| ~ | Around |
| % | Percentage |
| °C | Degree Celsius |
| µm | Micrometre |
| µL | Microliter |
| A | Adenine |
| AIC | Akaike information criterion |
| ap | Apicoplast |
| Ap5 | Apicoplast conserved tRNA region |
| BI | Bayesian Inference |
| BIC | Bayesian information criterion |
| BMC | Biological Species Concept |
| bp | Base pairs |
| C | Cytosine |
| CI | Confidence Interval |
| COI | Cytochrome c oxidase |
| Ct | Cycle threshold |
| Cytb | Cytochrome b |
| DNA | Deoxyribonucleic Acid |
| dNTP | Deoxynucleotide triphosphates |
| e.g. | Exempli gratia |
| et al. | Et alia |
| For (F) | Forward |
| *g* | Gravity units |
| G | Guanine |
| HAAR | Homopolymeric amino acid repeats |
| HMHZ | House Mice Hybrid Zone |
| ITS | Internal transcribed spacers |
| km | Kilometre |
| L/W | Length and Width ratio |
| maxEE | Maximum expected errors |
| MCMC | Markov Chains Monte Carlo |
| min | Minutes |
| mM | Millimolar |
| ML | Maximum Likelihood |
| MSC | Morphological Species Concept |
| mt | Mitochondria |
| *n* | Sample size |
| N | Any base |
| NCBI | National Center for Biotechnology Information |
| NMRI | Naval Medical Research Institute |
| nu | Nucleus |
| OPG | Oocysts per gram |
| ORF | Open reading frame |
| PCR | Polymerase Chain Reaction |
| PSC | Phylogenetic Species Concept |
| rRNA | Ribosomal Ribonucleic Acid |
| qPCR | Quantitative Polymerase Chain Reaction |
| Rev (R) | Reverse |

| | |
|---|---|
| s | Seconds |
| sp. | Species |
| spp. | Species pluralis |
| SNP | Single nucleotide polymorphisms |
| SRA | Short Read Archive |
| SSU | Small subunit |
| T | Thymine |
| tRNA | Transference Ribonucleic Acid |
| U/µm | Enzymatic Units per microliter |
| v. | Version |
| w/v | Weight per Volume |
| Y | Cytosine or Thymine |

# Chapter 1. General Introduction

## 1.1 Species and the morphology-phylogeny discrepancy in parasites

### 1.1.1 From characters to DNA sequences

Biology requires units to define, quantify and characterise the elements conforming the diversity of life on the earth. The unit employed to understand biodiversity, ecological and evolutionary processes is the species. How biologists understand and define this “unit of nature” has resulted in different approaches to conceptualize species. More than 20 different concepts of species have been established and are still employed[1]. However, the widespread use of different species concepts is a confounding influence describing biological diversity[2]. The probably most influential, the biological species concept (BSC) defines species as “groups of interbreeding natural populations that are reproductively isolated from other such groups”[3–5]. Despite the wide acceptance of the BSC among biologists, it has some drawbacks: the resulting criteria to delineate species are only applicable for sexual organisms and, it is not clear how can it be interpreted for asexual organisms or organisms with mechanism of horizontal gene transfer[6].

The use of interbreeding as a general criterion to delineate species is often impractical for pragmatic species identification. For that reason, specialists in systematics and taxonomy are indifferent to the BSC and tend to apply species status based on patterns of discrete variation[2]. Moreover, species identification approaches based on morphological or genetic characters are elevated to the level of concept, leading to a new era in the species-problem debate[7]. Considering that a universal species concept has yet to be established - if ever possible - , the present thesis will focus on morphological and phylogenetic identification approaches and their interrelations.

The morphological or Linnaean species concept (MSC) defines species as “the smallest group that is consistently and persistently distinct, and distinguishable by ordinary means”[8]. Taxonomists consider the MSC a reference for methods to identify species. Identifications account for diagnostic or unique characters that allow the distinction of members of one species from those of others. However, identification approaches rarely fulfil the “consistently and persistently” component of the concept[9,10]. Due to morphological plasticity, similar or even

identical characters could be found among unrelated organisms, but also organisms belonging to one species (according to e.g. the BSC) could display different traits in different environments[11–13]. A considerably trivial but still major concern regarding morphological species identification is the required expertise, which is often lost in modern biological sciences[14,15].

The phylogenetic species concept (PSC) defines species as “the smallest diagnosable cluster of individual organisms forming a monophyletic group within which there is a parental pattern of ancestry and descent”[16–19]. Thus, the PSC gives an evolutionary direction to the species concept, rather than a merely observational character. The cladistic nature of PSC can rely on well-established tree-based methods to delimit species. Inference of phylogenetic trees requires either genetic, or morphological characters or both[20]. However, genetic substitution patterns can be regarded as the most suitable proxy to identify members of a particular population, considering that a particular population carries the genetic inheritance from its ancestors[6]. Hence, the identification of species by the reconstruction of evolutionary history using genetic information as a methodological procedure and the phylogenetic species concept are in almost perfect accordance.

Technological advancements in nucleic acid extraction, DNA amplification and sequencing have favoured the generation of molecular data that has greatly benefitted phylogenetic reconstruction[21]. Species delimitation based on genetic information has disentangled and complemented species assignments based on morphological characters. Thus, phylogenetic analyses have accounted for discrepancies in previous classifications for some groups of eukaryotes[22–25]. Due to identification discrepancies, populations that do not form a single unique lineage lead to a radical reorganization of groups and rejection of “species” status. This situation reflects how the MSC and the BSC are rarely represented as robust identification methods compared to the PSC. Moreover, it has been possible to reveal genetically distinct populations that would otherwise be impossible to detect using only morphological data, particularly in the case of cryptic species[26]. Finding species that can be distinguished by one of the identification approaches but not the other hence generating numerous discrepancies, confusions and disagreement among taxonomists[27,28].

Despite the disagreement between morphological and phylogenetic species identifications, molecular taxonomists usually agree to incorporate the information provided by DNA sequences, as a tool to increase the efficiency of species identification[29–31]. DNA barcoding aims to overcome the challenges in taxonomy. This method links sequences from specific

marker DNA databases to ideally the type specimen broadly accepted as a member of a particular species[32]. Thus, a suitable DNA barcode should correspond to a given morphospecies, maximizing inter-species variation but simultaneously minimising the intra-species variability[33]. The use of DNA barcode sequences for phylogenetic analysis should, however, be interpreted with caution, as DNA barcodes can be insufficiently variable to uncover deeper phylogenetic relationships[34,35]. DNA barcodes and phylogenetic studies are broadly used as complementary strategies for the identification of problematic taxa. In particular, species lacking comprehensive information allowing the description by non-expert taxonomists are challenging to identify. Parasites represent an outstanding example of this situation[36–39].

### 1.1.2 Parasites, a taxonomic challenging group for morphology

Eukaryotic parasites are highly diverse and polyphyletic[40]. Animal parasites include ectoparasites and endoparasites, such as helminths and protozoans[41]. Therefore, the diversity of taxa included in the group of parasites requires different and even specific strategies of taxonomic assignment. Most of the assignments are strongly based on morphological traits. Moreover, a large number of species are only known by their original description and no further information is available to-date[42].

Host-mediated phenotypic variability represents an additional caveat of morphology-based parasite identification[43]. As a result, species diversity might be over- or underestimated. As an example, flagellates from the genus *Giardia* have an uncertain taxonomic classification based on the trophozoite morphology that originally led to the description of six species from different hosts. However, further DNA-based analysis grouped them in seven assemblages of genotypes within the species *G. duodenalis*[44,45]. In contrast, the helminth *Dipylidium caninum*[46] contains organisms with well-known host-induced morphological plasticity that are described as single species, despite genetic characterisation suggesting their separation into different species[47]. Hence, the identification of suitable molecular markers represents an essential task to overcome host-mediated phenotypic variability hindering morphological species identification.

### 1.1.3 Molecular markers for parasite phylogeny

In 2000, Carl Woese postulated the three domains of life and based his work on the small subunit rRNAs genes sequence analysis[48]. The tree of life represents a groundbreaking discovery for phylogenetic and evolutionary studies[49]. For some groups of parasites like

protists and helminths, the usage of ribosomal sequences as a molecular marker for phylogenetic inference and molecular systematics was already thoroughly used before Woese's work[50–53]. The amplification and sequence analysis of the nuclearly-encoded small subunit rRNA 18S (SSU 18S rRNA) has rapidly expanded, where its large-scale application has become a routine tool for molecular barcoding and phylogeny of Nematodes, Cestodes and Apicomplexa parasites[37,54,55]. Thus, SSU 18S rRNA has by far become the most frequently reported marker in reference databases not only for helminths, but also for a wide diversity of parasitic groups.

Ribosomal small subunit 18S rRNA phylogenetic analysis has successfully delineated parasite species and brought some conflict with morphological classifications[56–60]. Nevertheless, the lack of resolution or the influence of the concerted evolution within ribosomal loci, divergent paralogues, pseudogenes and recombinants, can sometimes lead to erroneous phylogenies[61–63]. Although SSU 18S rRNA gene has been proved to be extremely informative for resolving at higher taxonomic levels of Nematoda, for example[64], the relatively invariant sequences are not sufficient in all the cases for species-level discrimination[65]. Moreover, among members of the phylum Apicomplexa, the failure to achieve positional homology due to structural and functional heterogeneity of the SSU 18S rRNA represents the major reason of inconsistent or misleading phylogenies[66,67]. Thus, alternative genes are required to improve phylogenetic assessments.

Genes with high variability are usually selected when the aim is to infer phylogenies of organisms evolving over a short period[68]. On one hand, the internal transcribed spacers (ITS1 and ITS2) fulfil the criteria and have been used for phylogenetic reconstruction of *Neospora caninum,* as an example of parasites[69,70]. However, ITS sequences have the same disadvantages as SSU 18S rRNA linked to ribosomal cistron evolution restrict its usage. On the other hand, genes with mitochondrial or plastid origin that are protein-coding or non-coding have also been recommended in the phylogenetic analysis of parasites. Hence, cytochrome *c* oxidase subunit 1 (COI), and cytochrome *b* (cytb) were incorporated into the phylogenetic inference of apicomplexans and nematodes[71–73]. Moreover, the combination of genes in concatenated datasets provides higher resolution compared to individual genes phylogenetic inferences, as shown for different groups of parasites[74–76] including Coccidia[77,78]. Therefore, incorporating multiple genes should be generally encouraged to refine parasite species identification.

### 1.1.4 Multi-marker identification, an overlooked approach for parasites.

Multiple marker identification and multi-locus genotyping have been proposed as integrative

approaches to avoid the taxonomic misassignment based on a single DNA barcode[79]. These strategies can be either used to identify species or to study intra- and inter-species genetic variation. Multiple genetic marker identification has been used successfully to assess the diversity of prokaryotic and eukaryotic organisms[80], emphasizing the role of species discovery to aid the studies of host-parasite associations. Furthermore, the usage of multiple markers from different genomes (e.g. nuclear and mitochondrial) can unravel possible discordance between gene trees and species trees caused by incomplete lineage sorting that result in paraphyly and polyphyly of a subset of genes and hybridization[81]. However, multi-locus genotyping is not broadly applied to eukaryotic parasites, despite being an extensively used strategy for taxonomy and genetic characterization in prokaryotes[82,83]. Complexity of eukaryotic systems has deterred scientists from implementing this technique on parasites, which will be further addressed by this thesis. Studies conducted on particular protozoan and nematodes are an exception[84–88]. Thus, the implementation of multi-locus genotyping represents an unexplored but attractive strategy for other groups of parasites with debated taxonomy.

The different phylogenetic approaches for species identification, genealogy and population structure presented above have improved our comprehension on parasite taxonomy, but have consequently exposed the inconsistencies in the taxonomy of parasites. Hence, the discrepancies between phylogenetic and morphological classifications for different groups have become more evident. In this regard, Coccidia is a remarkable example of significant incongruences between morphology and phylogeny[78]. Besides, the integration of a broad range of phenotypic characters with their biology, life cycle, and host specificity has led to additional sources of incongruence within the phylogeny of Coccidia[89–91].

## 1.2 Host-parasite interaction and specificity

Parasites are, by definition, in continuous interaction with their host species and maintained at the expense of host resources[92]. From the perspective of parasites, these interactions imply a trade-off between mechanisms to resist the host defence strategies and to obtain resources without killing the host. As a result of this intimate interaction and mutual pressure, parasites might adapt either to a broad group of hosts or to specific ones[93]. The concepts of parasite specificity and generalism are based on their capacity to "use" one or multiple host species, respectively, at a given stage in the life cycle[94]. However, the degree of specificity is not an absolute characteristic or an all-or-nothing phenomenon. It represents a continuous scale transitioning from highly specialized parasites restricted to a single host, to more general ones

adapted to survive in multiple suitable hosts[95]. Throughout this thesis, host specificity will be primarily referred to, with the understanding that generalism represents the counter-part of this characteristic.

Host specificity is commonly measured as host range, despite being two different concepts. Host range is defined as the occurrence in which a determined parasite is found in a certain host(s)[96,97]. The host range assumes parasites have equal use to all host species and does not take into account the prevalence and intensity of the infection over different host genotypes and/or species[95,98] or the interaction with other organisms within the host[99]. Besides, evolutionary relationship or taxonomic distinctness among hosts used by a parasite represents a piece of complementary information to provide a comprehensive measure of host specificity[100].

The measure of host specificity could be biased by three aspects. First, an unequal sampling effort leading to repetitive records of a particular parasite in particular host species. For instance, studies in fish ectoparasites and ticks demonstrated a strong positive correlation between known host species and the frequency of reported parasite species[101–103]. Second, the adaptation of parasite populations to a locally available host is assumed in the estimation of host specificity for a given parasite species[104,105]. Hence, host specificity is directly linked to all members of a parasite species as a whole, which rarely takes interspecies variation into account[106]. Third, erroneous parasite species identification can provoke modifications to the host range assessment, resulting in either confusion or misinterpretation[107].

The impact of incorrect identification of parasites in host specificity measurement or even host range establishment could be overlooked and disregarded. The two following contrasting scenarios can arise: 1) Quantification of low host specificity for genetically different but morphologically indistinguishable organisms considered as a single species (Figure 1.1A) or 2) quantification of high host specificity for genetically indistinguishable but morphologically different organisms assigned as different species (Figure 1.1B). These two scenarios could simplify how parasite species identification represents a crucial step into accurate interpretations of host specificity[94,107].

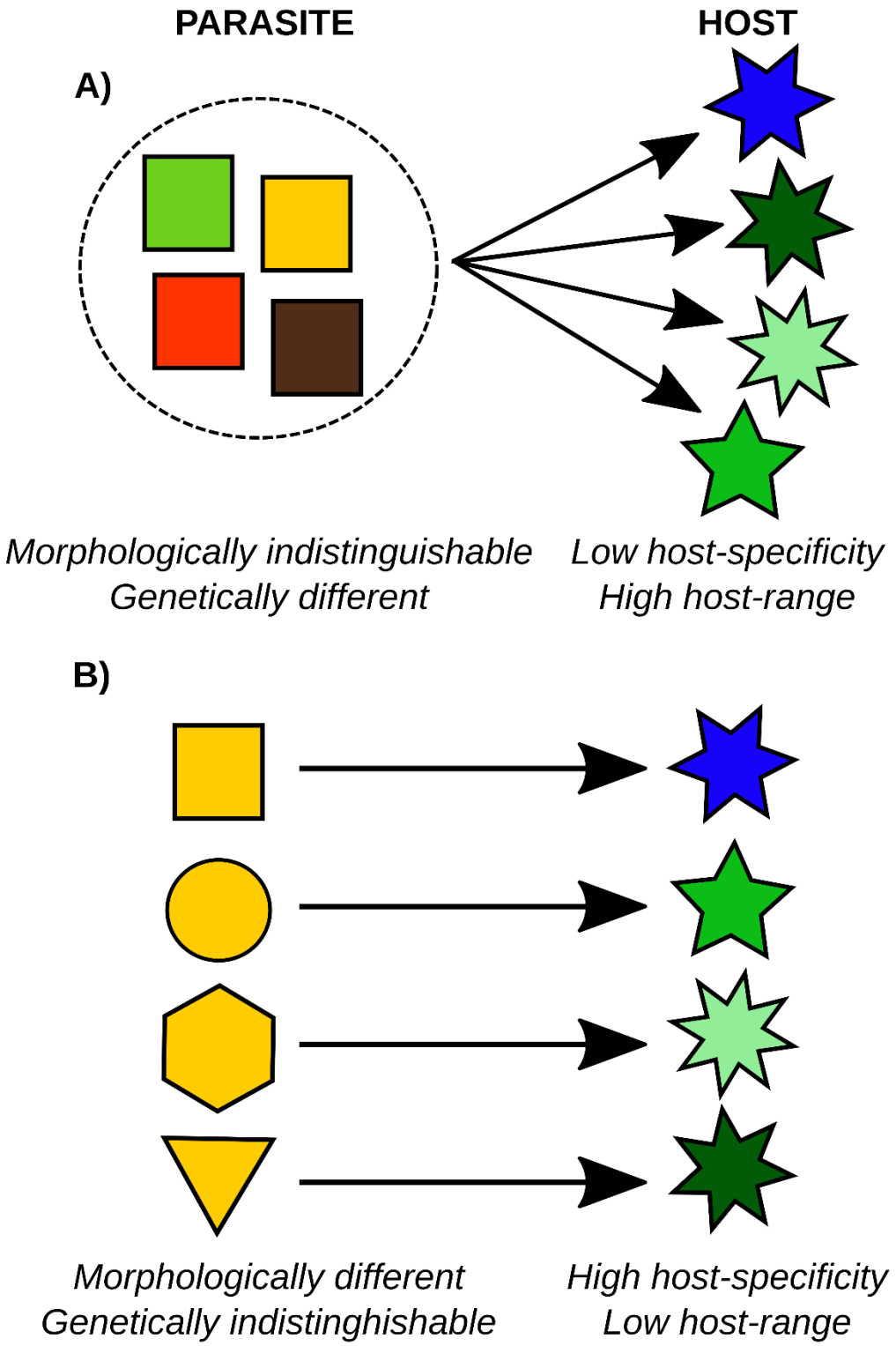



**Figure 1.1 Impact of parasite species identification on host specificity.** A) Assumption of low host specificity for genetically different but morphologically indistinguishable organisms considered as a single species. B) Assumption of high host specificity for genetically indistinguishable but morphologically different assigned as different species. For parasites and hosts: Shape indicates differences in morphology and colour indicates differences in genetics of the organism (Modified from Poulin and Keeney[107]).

A “vicious cycle” between host specificity and species identification is implied particularly when

the parasite-host defines the description of parasites and the number of hosts constitutes sufficient evidence for host specificity. In this regard, Coccidia parasites represent an ideal model to address this issue. Taken together, this thesis focuses on the morphological and the phylogenetic identification of coccidia from the genus *Eimeria*, while discussing the impact of host specificity in their classification.

## 1.3 Host-parasite system, *Eimeria* in rodents

### 1.3.1 Generalities of coccidians and *Eimeria*

Coccidia is a subclass under the phylum Apicomplexa. All organisms from this group are intracellular parasites characterised by the sporozoite and oocyst formation during life cycle progression. As with other members of the phylum, Coccidia has motility organelles only during the sexual microgamete stages. These parasites possess an apical complex and a vestigial plastid homologous to the chloroplasts of plants, known as apicoplast. The coccidian life cycle is composed of sexual (gametogony) and asexual (merogony) stages[108–110]. Despite taxonomic discrepancies, the so-called coccidia *sensu stricto* or class Eucoccidiorida includes the adeleid and eimerid protozoa. The latter includes parasites with clinical and veterinary relevance such as *Toxoplasma*, *Sarcocystis, Isospora, Cyclospora* and *Eimeria*[111].

*Eimeria* is the largest genus of the phylum Apicomplexa (~1,700 species), and infects a wide variety of vertebrate hosts, including birds, mammals, some reptiles and even fish[112–116]. It comprises intracellular homoxenous protozoan parasites that infect host cells of the gastrointestinal tract. Unsporulated oocysts are formed during the sexual stage of the parasite, leaving the host as an undeveloped and thus non-infective stage. In the external environment, under optimal oxygen, moisture and temperature conditions, unsporulated oocysts mature and form infective sporulated oocyst that contains four sporocysts with two haploid sporozoites each[117]. *Eimeria* spp. have been usually described as niche- and host-specific[118], although host switching has been demonstrated over evolutionary time[90]. The infective stages of *Eimeria* are transmitted through faecal-oral route. Even though infections are generally mild and self-limiting, and might produce resistance against re-infection[119,120], certain *Eimeria* species are well-known to cause severe infections in livestock animals. This represents a high economic burden in poultry production annually[121,122].

### 1.3.2 Species descriptions and phylogeny in *Eimeria*

Historically, identification of *Eimeria* parasites has been based on the morphology and morphometry of sporulated oocysts[112,113,115]. Different qualitative (e.g. the outer oocyst wall texture and presence/absence of micropyle or Stieda body) and quantitative morphological features (e.g. the shape index) were essential for original descriptions and have been the reference for later identifications[123]. However, most of the original *Eimeria* records lack sufficient details, rendering them unreliable to a certain extent and the subsequent identification process becomes problematic[124,125]. At best, the descriptions are accompanied by schemes of the diagnostic structures, which may be confusing or misleading, especially for inexperienced parasitologists. Furthermore, reported polymorphism of oocyst increases the challenges in morphological identification[126,127]. Therefore, the systematic record of host species became indispensable as part of the guidelines for species description, intending to clarify the host specificity of *Eimeria* parasites as an important criterion for identification[128,129].

Molecular analysis of *Eimeria* has benefited from the presence of additional extrachromosomal genomes, one in the mitochondria and one in the apicoplast. Therefore, amplification and sequence regions from the three genomes might provide better phylogenetic resolution[72]. Although previous studies have suggested an association between phylogenetic diversity and the oocyst residuum[130–132], phylogenetic inference of multiple genes confirmed that *Eimeria* general oocyst morphology and morphotypes do not correlate with the phylogeny[77,78]. Collectively, these findings might lead to a radical taxonomic re-arrangement of the genus *Eimeria*. Moreover, the general idea of host-parasite coevolution and specificity is still supported, even though phylogenetic analysis found incongruences in the clustering of *Eimeria* species from different hosts[90,132]. Further studies have tried to explain the inconsistencies by the presence of differentiation in the parasite population, based on the ecology of the host[91]. Thus, the implementation of multi-marker approaches could be beneficial to solve the inconsistencies regarding the genetic diversity of *Eimeria* and the relationship to their host.

### 1.3.3 Rodent-infecting *Eimeria*

Rodent-infecting *Eimeria* corresponds to the vast majority of species already described for this genus, including 374 named species and 32 additional ones with incomplete information to justify species assignment[112,133]. These described species are distributed among 5 families of rodents: Muridae, Cricetidae, Heteromyidae, Bathyergidae and Gliridae, representing less than 15% over the total diversity of rodents described so far[134].

Phylogenetically, rodent-infecting *Eimeria* clusters separately and form at least three independent lineages. The first one contains *Eimeria* species infecting endemic North and Central American Heteromyidae and Muridae rodents[135,136]. The second encloses most of the

species described in Muridae hosts *Mus musculus* and *Rattus* (e.g. *E. falciformis*, *E. ferrisi*, *E. vermiformis*) but also in non-*Mus* Muridae from the genus *Apodemus*. The third consist of *E. myoxi* described in *Eliomys quercinus* and defined as "squirrel-related host clade"[131,137]. However, the incorporation in recent phylogenies of new isolates from non-*Mus* murids and cricetid rodents suggests a larger number of lineages[90,91].

Within the Muridae family, up to 16 species of *Eimeria* have been described from house mice (*Mus musculus*). *Eimeria falciformis* was the first species described[138,139], and therefore, became the type species for rodents. The adaptation of *E. falciformis* to different laboratory mice strains have facilitated the study of host-parasite interaction at immunological and metabolic mechanism level[140,141], as well as parasite interaction with gut microbiota[142]. Moreover, laboratory adaptation of this parasite led to the sequencing of the only publicly available genome of a rodent-infecting *Eimeria*[143]. However, recent studies in our group indicated the different life-cycle, virulence patterns and host resistance induced by different *Eimeria* species in wild-derived and laboratory mice from the same species[144,145]. These results suggest that observations of a single *Eimeria* species cannot be generalized for all *Eimeria* species infecting different hosts from a diverse group such as rodents. Hence, *Eimeria* species identification within a single host species system (**Chapter 2**) becomes relevant and the stepstone to understanding the evolutionary and epidemiological role of these parasites in wild rodent populations.

The host specificity of rodent-infecting *Eimeria* is critical for species identification. It is suggested that rodent-infecting *Eimeria* do not cross family or even genus boundaries except for *E. chinchillae*[146]. This assumption requires the implementation of cross-infection experiments as definitive evidence of transmission among different host species[147–149]. However, cross-infection experiments are not always feasible for wild-derived isolates due to insufficient or limited material for infection, and hence the approach becomes unviable. Phylogenetic analysis suggests that the host specificity might not represent the most important role in *Eimeria* species clustering[90,150]. Therefore, the implementation of novel and high-sensitive molecular methodologies is fundamental for differentiating wild isolates from diverse host species (**Chapter 3**).

## 1.4 Aims of the thesis

The ultimate goal of this thesis is the development and refinement of approaches for parasite identification. Specifically, how these approaches can be combined to (1) describe parasite diversity in natural systems and (2) understand how the species identification is linked to host-parasite interaction characteristics, especially host range of parasites. Rodent coccidia from the genus *Eimeria* are used as the study model. To achieve the general goal, two specific aims were established:

1) ***To provide methods for the assessment of coccidia prevalence at the species level in rodent systems.***

In Chapter 2, *Eimeria* species identification and quantification in a single host species system are examined. The prevalence of these parasites are assessed in free-living populations of house mice (*Mus musculus*) from a transect within the European house mice hybrid zone (HMHZ). Conventional morphological methods in conjunction with sequence analysis and phylogenetic inference with established makers are implemented in these wild commensal populations of house mice.

2) ***To assess the diversity of Eimeria isolates from wild rodents at shallow depths of phylogenetic relationships.***

Chapter 3 explores the use of multi-marker phylogenetics to discuss whether host association is a relevant factor to species identification of morphologically indistinguishable *Eimeria* isolates from different host species. Host specificity is first evaluated based on phylogenetic analysis using established markers from the nuclear, mitochondrial and apicoplast genome. Then, to test how far these markers are polymorphic enough to resolve between genetic clusters with different host usage, a multi-locus sequence typing method is applied.

# Chapter 2. Detection and quantification of house mouse *Eimeria* at the species level - challenges and solutions for the assessment of Coccidia in wildlife

**Published article**

Jarquín-Díaz, V. H., Balard, A., Jost, J., Kraft, J., Dikmen, M. N., Kvičerová, J., & Heitlinger, E.

*International journal for parasitology. Parasites and wildlife* (2019), *10*, 29–40.

https://doi.org/10.1016/j.ijppaw.2019.07.004



**Author contributions:**

VHJD and EH designed the project and obtained funding. VHJD, AB, JJ, JK and MND obtained data, VHJD led the data analysis and interpreted the results with support of EH, AB and JK. VHJD and EH wrote the manuscript with contributions from all other authors. EH supervised the project.

## 2.1 Abstract

Detection and quantification of coccidia in studies of wildlife can be challenging. Therefore, prevalence of coccidia is often not assessed at the parasite species level in non-livestock animals. Parasite species – specific prevalences are especially important when studying evolutionary questions in wild populations. We tested whether increased host population density increases prevalence of individual *Eimeria* species at the farm level, as predicted by epidemiological theory.

We studied free-living commensal populations of the house mouse (*Mus musculus*) in Germany, and established a strategy to detect and quantify *Eimeria* infections. We show that a novel diagnostic primer targeting the apicoplast genome (Ap5) and coprological assessment after flotation provide complementary detection results increasing sensitivity. Genotyping PCRs confirm detection in a subset of samples and cross-validation of different PCR markers does not indicate bias towards a particular parasite species in genotyping. We were able to detect double infections and to determine the preferred niche of each parasite species along the distal-proximal axis of the intestine. Parasite genotyping from tissue samples provides additional indication for the absence of species bias in genotyping amplifications. Three *Eimeria* species were found infecting house mice at different prevalences: *Eimeria ferrisi* (16.7%; 95% CI 13.2 – 20.7), *E. falciformis* (4.2%; 95% CI 2.6 – 6.8) and *E. vermiformis* (1.9%; 95% CI 0.9 – 3.8). We also find that mice in dense populations are more likely to be infected with *E. falciformis* and *E. ferrisi*.

We provide methods for the assessment of prevalences of coccidia at the species level in rodent systems. We show and discuss how such data can help to test hypotheses in ecology, evolution and epidemiology on a species level.

**Key words**

*Eimeria*, Coccidia, house mice, diagnostic PCR, species-specific prevalence, qPCR

## 2.2 Introduction

House mice (*Mus musculus*) are the most commonly used mammalian model for biomedical research worldwide[151,152]. Laboratory mouse strains are derived mainly from the subspecies *M. m. domesticus* with genetic contributions from other subspecies (*M.m musculus* and *M. m. castaneus*)[153,154]. Establishment of suitable mouse models to better understand infections with

coccidia is an ongoing process. Wild rodents and especially wild house mice are an attractive system for first steps in this direction[155].

*Eimeria*[139] is, with around 1,700 species, the most speciose genus in the phylum Apicomplexa[117,156]. For economic reasons, the most studied parasites in this group are those infecting livestock[157,158]. At least one third of the described species, however, infects rodents[112,159].

The most commonly used method for detection and identification of coccidia is the flotation and microscopical observation of oocysts shed in faeces during the patent period of infection[160]. Unsporulated oocysts, however, are difficult or impossible to differentiate into species[112,147]. Thus, prior to identification the oocyst should be sporulated under specific conditions. In addition, expertise and experience is required for species identification, especially in cases (like ours) of very similar oocyst morphology in different species[124,129]. For that reason, tools based on DNA amplification and sequencing have been included as routine strategy not only for detection, but also for taxonomic assessment[158,161–166].

Up to 16 species of *Eimeria* have been described from house mice[112] and some of them use different niches in the intestine. The reasons for this diversity are still elusive[131] and artificial splitting of morphologically plastic forms of the same species (in the same of different hosts) might contribute to this.

*Eimeria* species described from house mice include *E. falciformis,* the first coccidia described in house mice[138], which has sometimes been regarded as the most prevalent species in mice[108,167]. This species (and especially the BayerHaberkorn1970 isolate) are the most commonly studied coccidia model in laboratory mice. Life cycle progression[168] and host response[140,169,170] are relatively well studied and the whole genome of this species has been sequenced and annotated in detail[143].

*E. vermiformis* was first described in 1971[171] but since then, to our knowledge, not reported in wild house mice. Similar to *E. falciformis*, most of the information on this species comes from laboratory infection experiments[172–175], making the timing of life cycle progression and its effect on the host relatively well studied.

*E. ferrisi* was originally described from *M. m. domesticus* from North America[112,176]. Laboratory infections with this parasite have confirmed its shorter life cycle, compared to *E. vermiformis* or *E. falciformis*[177].

To the best of our knowledge, just few investigation of prevalences and intensities of Coccidia

has been conducted in free-living populations of *M. musculus*[171,178–182].

In the present work we studied the prevalence of *Eimeria* in house mice from a transect of the well-studied European house mice hybrid zone (HMHZ)[183–185]. We established methods for detection, species identification and quantification of *Eimeria* in these wild commensal populations of house mice.

## 2.3 Material and methods

### 2.3.1 Collection of samples

Between 2015 and 2017, 378 house mice (*Mus musculus*) were captured in 96 farms and private properties in a transect 152.27 km long and 114.48 km wide, within the German federal state of Brandenburg (capture permit No. 2347/35/2014) (Supplementary data S2.1). On average 20 traps were set overnight per locality. Mice were housed individually in cages overnight and euthanised by cervical dislocation. All mice were dissected within 24 hours after capture. Faeces for microscopical diagnosis of *Eimeria* spp. were preserved in potassium dichromate ($K_2Cr_2O_7$) 2.5% (w/v) and stored at 4°C until further processing, colon content was preserved in liquid nitrogen and stored at -80°C. For a subset of 163 mice (from Brandenburg in 2016) tissue samples from cecum and ileum were collected for DNA extraction and molecular identification of *Eimeria* spp. All samples were kept in liquid nitrogen during transportation and maintained at -80°C until processing.

### 2.3.2 Flotation and microscopical analysis of oocyst

Fecal samples were washed with tap water to eliminate potassium dichromate and homogenized. After oocyst were floated using a saturated salt solution (specific gravity = 1.18—1.20), they were collected by centrifugation (3,234 × *g*/room temperature/ 10 minutes), washed with distilled water (1,800 × *g*/room temperature/10 minutes). The flotations were screened for the presence of oocyst using a Leica® DM750 M light microscope under the 10X objective. To estimate the intensity of infection, floated oocysts were counted using a Neubauer chamber and the results were expressed as oocyst per gram (OPG) of faeces. Samples were then preserved in a fresh solution of potassium dichromate 2.5% (w/v) and sporulated in a water bath at 30°C for 10-12 days for further characterisation.

*Eimeria* isolates, corresponding to different phylogenetic groups (see below), were selected to take photomicrographs of sporulated oocysts using a Carl-Zeiss microscope at 100x

magnification. Measurements were made on ~30 oocysts and ~30 sporocysts, using Adobe Photoshop CC v14.2.1 (3778 pixels = 100, 000 µm). Length and width were measured and reported in micrometers. The Length/Width (L/W) ratio was calculated for both oocysts and sporocysts including means, standard deviation and variation coefficients. Additionally, main morphological traits (oocyst wall, oocyst residuum, micropyle, polar granule, sporocyst residuum, refractile bodies and Stieda body) were described, according to the protocol of Duszynski and Wilber (1997)[129].

### 2.3.3 DNA extraction

DNA from colon content was extracted using the NucleoSpin® Soil kit (MACHEREY-NAGEL GmbH & Co. KG, Düren, Germany) according to the manufacturer's recommendations, adding a mechanical lysis process in a Mill Benchtop Mixer MM 2000 (Retsch GmbH, Haan, Germany). DNA from cecum and ileum tissues was isolated using the innuPREP DNA Mini Kit (Analytik Jena AG, Jena, Germany) following the instructions of the manufacturer after disruption of the tissue with liquid nitrogen in a mortar. Quality and quantity of isolated DNA was measured spectrophotometrically in a NanoDrop 2000c (Thermo Scientific, Waltham, USA).

### 2.3.4 PCR amplification for detection (ap tRNA) and identification (nu 18S rRNA and mt COI)

For detection of *Eimeria*, amplification of a conserved tRNA region of the apicoplast genome (Ap5) was used. Primers Ap5_Fwd (YAAAGGAATTTGAATCCTCGTTT) and Ap5_Rev (YAGAATTGATGCCTGAGYGGTC) were designed based on the complete apicoplast genomes sequences available in the GenBank from *Eimeria tenella* (NC_004823.1), *E. falciformis* (CM008276.1) and *E. nieschulzi* (JRZD00000000.1).

For all samples with oocysts detected during flotation or successful amplification of Ap5, genotyping PCRs were performed to confirm detection and further identification of parasite species. A fragment of nuclear small subunit ribosomal RNA (~1,500 bp) and mitochondrial cytochrome *C* oxidase subunit I (~800 bp) were amplified using primers 18S_EF and 18S_ER[132] and Cocci_COI_For/Rev[72], respectively. An alternative pair of primers was used in case of failure to amplify COI: Eim_COI_M_F (ATGTCACTNTCTCCAACCTCAGT) and Eim_COI_M_R (GAGCAACATCAANAGCAGTGT). These primers were designed based on the mitochondrial genome of *E. falciformis* (CM008276.1)[143] and amplify a ~700 bp fragment of COI gene.

PCR reactions were carried out in a Labcycler (SensoQuest GmbH, Göttingen, Germany) using 0.025 U/µL of DreamTaq™ DNA Polymerase (Thermo Scientific, Waltham, USA), 1X DreamTaq Buffer, 0.5 mM dNTP Mix, 0.25 µM from each primer and 1—20 ng/µL of DNA template in 25 µL reaction. A concentration of 0.25 mM dNTP Mix and a supplementation with 2 mM $MgCl_2$ was used for the amplification of Ap5. The thermocycling protocol consist of 95 °C initial denaturation (4 min) followed by 35 cycles of 92 °C denaturation (45 s), annealing at 52°C (30 s/Eim_COI); 53 °C (45 s/18S_E); 55 °C (30 s/Cocci_COI); 56 °C (30 s/Ap5); 72 °C extension 90 s (18S_E), 20 s (Cocci_COI/Eim_COI) or 45s (Ap5) and a final extension at 72 °C (10 min). DNA from oocyst of *E. falciformis* BayerHaberkorn1970 strain and DNA from colon content of a non-infected NMRI mouse were used as positive and negative controls, respectively.

All PCR products with the expected size were purified using the SAP-Exo Kit (Jena Bioscience GmbH, Jena, Germany) and Sanger sequenced from both directions by LGC Genomics (Berlin, Germany). Quality assessment and assembly of forward and reverse sequence was performed in Geneious v6.1.8. All sequences were submitted to NCBI GenBank (Accession numbers: nu SSU 18S rRNA [MH751925—MH752036] and mt COI [MH777467—MH777593 and MH755302—MH755324] (Supplementary data S2.2).

### 2.3.5 *Eimeria* detection in tissue by qPCR

For mice collected in 2016 (*n*= 163) cecum and ileum tissue was screened using qPCR. Primers targeting a short fragment of mt COI were used to amplify DNA from from intracellular stages of *Eimeria* (Eim_COI_qX-F, TGTCTATTCACTTGGGCTATTGT; Eim_COI_qX-R GGATCACCGTTAAATGAGGCA). Amplification reactions with a final volume of 20 µL contained 1X iTaq™ Universal SYBR® Green Supermix (Bio-Rad Laboratories GmbH, München, Germany), 400 nM of each primer and 50 ng of DNA template. Cycling in a Mastercycler® RealPlex 2 (Eppendorf, Hamburg, Germany) was performed with the following program: 95 °C initial denaturation for 2 min, followed by 40 cycles of denaturation at 95 °C for 15 s, annealing at 55 °C for 15 s and extension 68 °C for 20 s. Melting curves were analysed to detect eventual primer dimer formation or non-specific amplification. As internal reference for relative quantification the CDC42 gene from the nuclear genome of the house mouse was amplified (Ms_gDNA_CDC42_F CTCTCCTCCCCTCTGTCTTG; Ms_gDNA_CDC42_R TCCTTTTGGGTTGAGTTTCC). Infection intensity was estimated as the ΔCt between mouse and *Eimeria* amplification (CtMouse- Ct*Eimeria*). To correct for background noise a detection threshold was estimated at ΔCt = -5 and only results above this value were considered infected.

ΔCtIleum and ΔCtCecum were compared for samples above the threshold in both tissues to assess primary tissue occurrence[186]. In samples positive for qPCR, *Eimeria* genotyping was performed based on DNA extracted from tissue, as described above (see 2.3.4).

### 2.3.6 Molecular identification of *Eimeria* spp. isolates: 18S and COI phylogenetic analysis.

As a strategy for molecular identification, datasets of nu 18S and mt COI sequences were compiled. Sequences generated for the present work were compared to database sequences using NCBI BLAST and most similar sequences were selected. Based on this, sequences for *E. falciformis, E. vermiformis* and *E. ferrisi* were downloaded from GenBank as a reference. COI sequences were aligned by translation using the Multiple Align algorithm and translation frame 1 with the genetic code for "mold protozoan mitochondrial", 18S sequences were aligned using MUSCLE[187], both through Geneious v6.1.8.

Phylogenetic trees for all datasets were constructed using Maximum Likelihood (ML) and Bayesian inference (BI) methods, implemented in PhyML v3.0[188] and MrBayes v3.2.6[189,190], respectively. The most appropriate evolutive models for each dataset were determined in JModelTest v2.1.10[191]. For ML trees, a bootstrap analysis with 1000 replicates was performed, whereas MCMC for BI was run with two cold and two hot chains for 1,000,000 generations or until the average split frequency was below 0.05. The concatenated dataset was analysed using partitions and locus-specific models. Visualization of the trees was done with FigTree v1.4.2[192].

### 2.3.7 Statistical analysis

All statistical analyses were performed in R (R Development Core Team, 2008). Prevalence of *Eimeria* was calculated as the proportion of positive samples in the total number of analysed samples. The 95% confidence interval [CI 95%] was calculated using the Sterne's exact method[193] implemented in the package "epiR" v0.9-99[194]. Prevalences were tested for statistical differences with the Fisher's exact test [195].

To assess the significance in primer bias, logistic regression models were used to estimate the probability to successfully amplify and sequence a specific genetic marker for each *Eimeria* species. The response variable in these models was the amplification and sequencing success with a particular primer pair (COCCI_COI_F/R, Eim_COI_M_F/R or 18S_E F/R), and the predictors were the species identity (as determined with the other markers only, to make

response and predictors independent) and additionally the detection of an infection with Ap5 and Flotation. These models were fitted first for COCCI_COI_F/R as response, then for the combined probability of successful COI genotyping and finally for 18S genotyping as response. Tables were produced for the summary of models using the package "jtools" v2.0.0[196].

For each *Eimeria* species logistic regression models were used to test whether the infection is influenced by host density or by the presence of the other two *Eimeria*. We use as response variable the infection status by *E. ferrisi*, *E. falciformis* or *E. vermiformis*, independently, and the total number of mice cough per locality per year and the infection status by a different *Eimeria* species as predictors.

Differences on oocyst and sporocyst L/W ratios between *Eimeria* species were tested for significance with an analysis of variance fitting a linear model using the species assignment as predictor with a Tukey HSD post hoc test adjusting for multiple comparisons.

## 2.4 Results

### 2.4.1 Sampling and *Eimeria* spp. detection

We used flotation of oocyst from faeces and PCR amplification of a novel diagnostic maker (Ap5) from colon content DNA to detect *Eimeria* parasites in a total of 378 house mice. Overall prevalence was 25.9% [95% CI = 21.7 – 30.7] (98/378) for PCR and 27.0% [95% CI = 22.7 – 31.7%] (102/378) for flotation. These estimates are not significantly different (Fisher exact test, $p > 0.05$). However, both techniques considered together estimate a higher prevalence of 37.6% [95% CI = 32.8 – 42.6] (142/378), meaning that 44 and 40 positive results were detected only by flotation or PCR, respectively (Figure 2.1). We further aimed to provide species specific identification and to consolidate results from the two different detection methods.

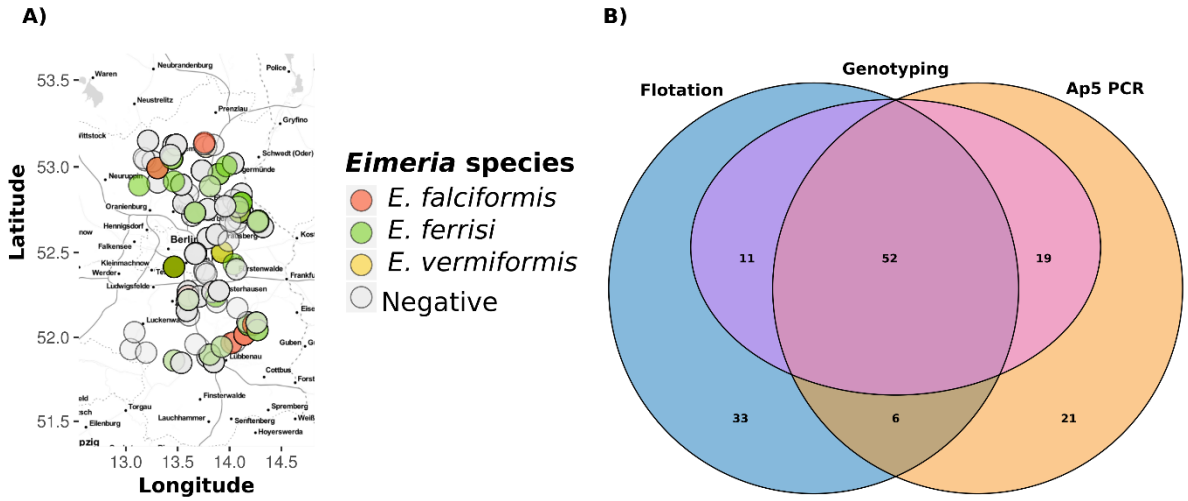



**Figure 2.1 Geographical localization of house mice (*Mus musculus*) collected for this study and comparison of diagnostic methods for *Eimeria*.** A) Localization from the 378 mice included in the present study, colors indicate the *Eimeria* species identified for each. B) Venn diagram showing the overlap between detection methods and successful genotyping identification of the isolates.

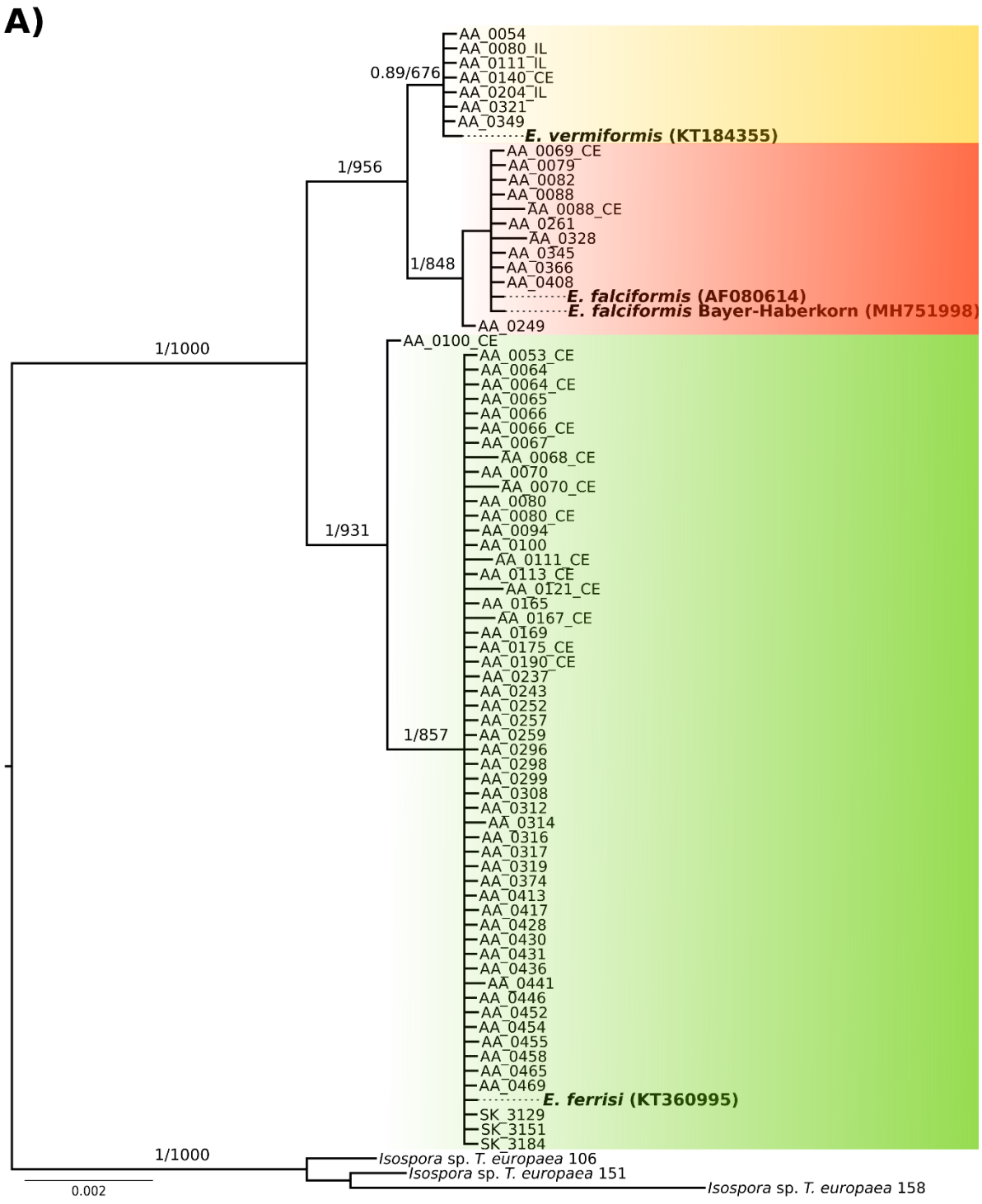
### 2.4.2 Molecular identification of *Eimeria* isolates - (phylogenetic analysis nu 18S and mt COI)

*Eimeria* species were identified by phylogenetic analysis of nu 18S and mt COI sequences, the most commonly used molecular markers of apicomplexan parasites. To identify our isolates, sequences were compared with references from the NCBI database. Sequences from three previously described *Eimeria* species infecting *M. musculus* showed highest BLAST similarities and phylogenetic clustering. This approach ignores the problem of whether isolates from different hosts would be assigned to the same phylogenetic clusters while they are regarded as different species by taxonomists.

The nu 18S phylogenetic tree was inferred based on 80 sequences (540—1,795 bp), 73 of them from wild house mice generated in our study (3 from ileum tissue, 16 from cecum tissue and 54 from colon content, see below). *Eimeria* species previously described in house mice were represented by *E. falciformis* (AF080614), *E. vermiformis* (KT184355) and *E. ferrisi* (KT360995). In addition, one newly generated sequence from *E. falciformis* strain BayerHaberkorn1970 (MH751998) was also included. Sequence identity of our isolates to these reference sequences was above 98% and even 100% in most of the cases for this marker. *Isospora* sp. sequences identified in *Talpa europaea* moles were used as outgroup. Both ML and BI rooted trees shared a topology placing our sequences at the same positions in relation to reference sequences with high support (bootstrap values and posterior

probabilities are shown in Figure 2.2A). The sequences clustered in three well supported monophyletic groups (Figure 2.2A).

The phylogenetic tree for mt COI was based on 103 sequences (519—804 bp), 97 of which were obtained from *Eimeria* infecting wild house mice (3 from ileum, 16 from cecum tissue and 78 from colon content). Reference sequences from house mouse *Eimeria* (*E. ferrisi*, KT361028; *E. falciformis,* KX495129 and MH777539; *E. vermiformis* JN205071) identified by BLAST searches showed an identity of above 98% to respective groups of our isolates. We defined *Isospora* sp. from *Talpa europaea* as an outgroup for rooting. ML and BI rooted trees based on alignments of these COI sequences shared a general topology with respect to the placement of our isolates in relation to reference sequences. Bootstrap values and posterior probabilities for support of these placements are shown in Fig. 2.2B.




Continue next page

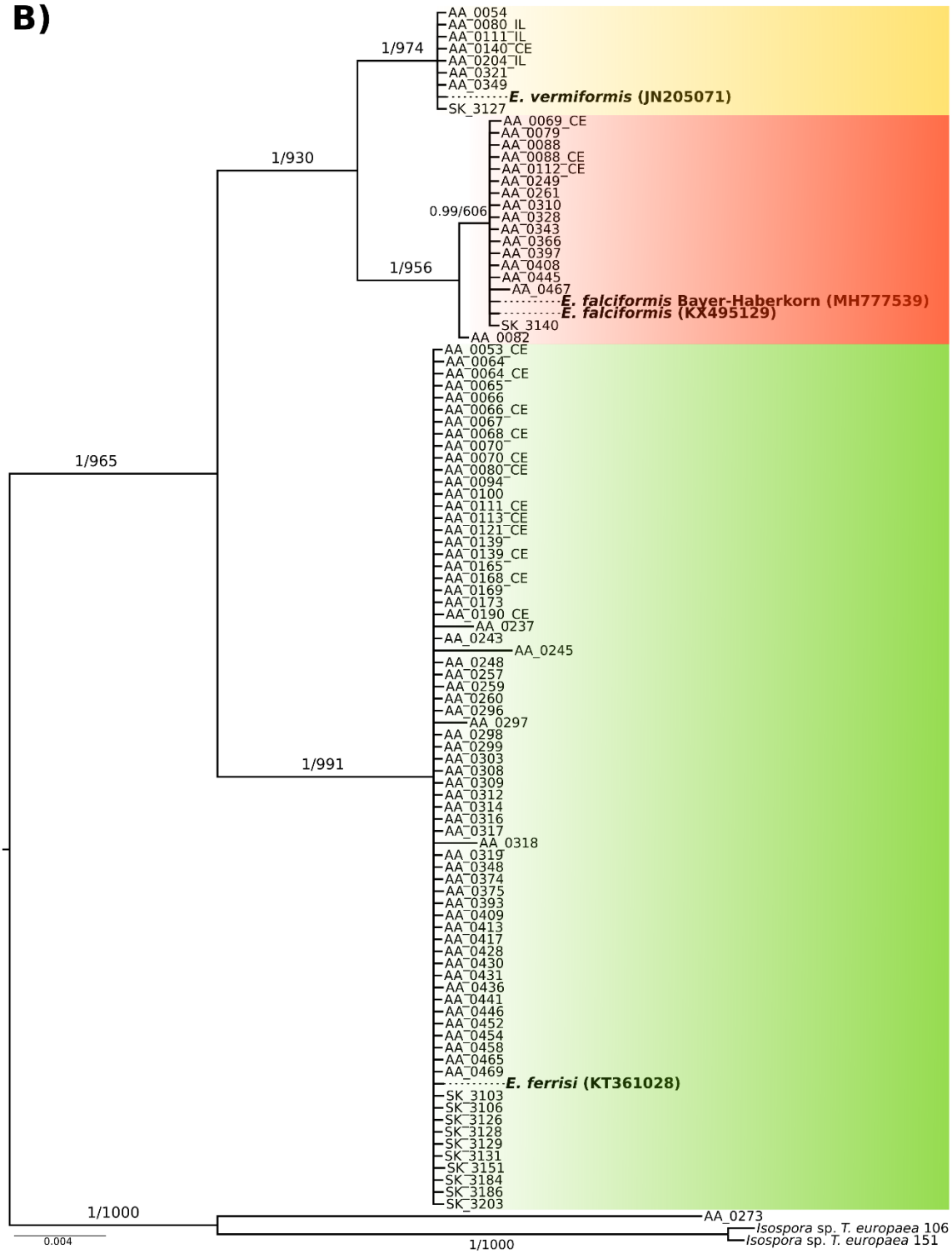



**Figure 2.2 Phylogenetic trees based on 18S rRNA and COI sequences.** Sequences of 18S A) and COI B) were used to infer the molecular identification of wild-derived isolates of *Eimeria*. In both phylogenies, our isolates are clustered in three groups one close to *E. falciformis* (red), other close to *E. ferrisi* (green) and finally one to *E. vermiformis* (yellow). Numbers in the branches represent the Bayesian posterior probability and the non-parametric bootstrap value. In bold are the reference sequences for each species. CE and IL make reference to sequences derived from cecum or ileum tissue DNA, respectively.

The sequences derived from house mice cluster in three monophyletic groups including reference sequences for *E. falciformis* (*n*= 17, sequences from our study), *E. ferrisi* (*n*= 72) and *E. vermiformis* (*n*= 8) (Figure 2.2B). Phylogenies based on concatenated supermatrices for the two markers show the same topology concerning placement of isolates from the present study (Supplementary data S2.3 and S2.4).

### 2.4.3 Morphometrical and morphological comparison of oocysts

For further support assignment of the three phylogenetic groups of *Eimeria* from house mouse, we characterized sporulated oocysts morphologically (Table 2.1). *E. falciformis, E. ferrisi* and *E. vermiformis* oocyst shared most of the traits we evaluated and showed overlapping morphometry (Figure 2.3A). The length/width ratio of *E. vermiformis* oocysts, however, was significantly higher (1.29; 95%CI = 1.26—1.33; *n*= 35) than that of *E. falciformis* (1.17; 95%CI = 1.14—1.20; *n*= 31) and *E. ferrisi* oocysts (1.23; 95%CI = 1.21—1.25; *n*= 127) (Tukey HSD, $p < 0.05$) (Figure 2.3B; Supplementary data S2.5). This means that *E. vermiformis* has more ellipsoidal oocysts than the other two species. Other morphological characteristics of oocysts (smooth wall, absence of micropyle, presence of polar granule and absence of oocyst residuum) are very similar or identical between the three species (Table 2.1).

Morphological measurements of sporocysts are not significantly different between the three species found in house mice (Fig. 2.3C). We also observed the presence of a sporocyst residuum, refractile bodies and Stieda bodies in all species uniformly, in agreement with previous descriptions[138,168,171,176].

**Table 2.1 Morphological and morphometrical characteristics from *Eimeria* wild-derived isolates and reference morphotypes[§].**

| Species | Oocysts | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Shape | Wall | Micropyle | Polar granule | Residuum | Lenght (µm) | Width (µm) | L/W ratio |
| *E. falciformis* | Spherical/Ellipsoidal | Smooth | Absent | Present | Absent | 18.62 (15.29 – 20.76) | 15.92 (13.89 – 17.6) | 1.17 (1.00 – 1.36) |
| *E. falciformis* | Ovoid/Spherical | Smooth | Absent | Present | Absent/Present | 21 (15 – 26) | 18 (13 – 24) | 1.17 (1.1 – 1.2) |
| *E. ferrisi* | Spherical/Ellipsoidal | Smooth | Absent | Present* | Absent | 17.32 (13.59 – 21.47) | 14 (10.97 – 17.44) | 1.23 (1.02 – 1.56) |
| *E. ferrisi* | Ellipsoidal/Subspherical | Smooth | Absent | Present | Absent | 17 (12 – 22) | 14 (11 – 18) | 1.22 (1.0 – 1.6) |
| *E. vermiformis* | Spherical/Ellipsoidal | Smooth | Absent | Present* | Absent | 20.02 (16.22 – 22.8) | 15.57 (12.37 – 18.44) | 1.29 (0.99 – 1.45) |
| *E. vermiformis* | Spherical/Ellipsoidal | Lightly pitted | Absent | Present | Absent | 23.1 (18 – 26) | 18.4 (15 – 21) | 1.25 (1.1 – 1.4) |

| Species | Sporocysts | | | | | Tissue localization | Reference |
|---|---|---|---|---|---|---|---|
| | Residuum | Refractile body | Stieda body | Lenght (µm) | Width (µm) | | |
| *E. falciformis* | Present | Present | Present** | 8.88 (7.03 – 10.30) | 5.52 (4.16 – 6.93) | Cecum | This work |
| *E. falciformis* | Present | Present | Present | 11 (10 – 12) | 7 (6 – 8) | Ileum,**cecum** and colon | Eimer, 1870; Haberkorn, 1970 |
| *E. ferrisi* | Present | Present | Present | 7.75 (5.17 – 11.47) | 5.03 (2.88 – 7.73) | Cecum | This work |
| *E. ferrisi* | Absent/Present | Present | Present | 10.5 (10 – 11) | 5.5 (5 – 6) | **Cecum** and colon | Levine & Ivens, 1965; Ankrom et al., 1975 |
| *E. vermiformis* | Present | Present | Present | 8.35 (6.18 – 11.29) | 5.43 (4.19 – 7.00) | Cecum and ileum | This work |
| *E. vermiformis* | Present | Present | Present | 12.8 (11 – 14) | 7.9 (6 – 10) | Jejunum,**ileum** and cecum | Ernst et al., 1971 |

§ Measurements are means in micrometers with ranges in parenthesis
*Observed in more than 80% of the oocysts
**Present but not evident

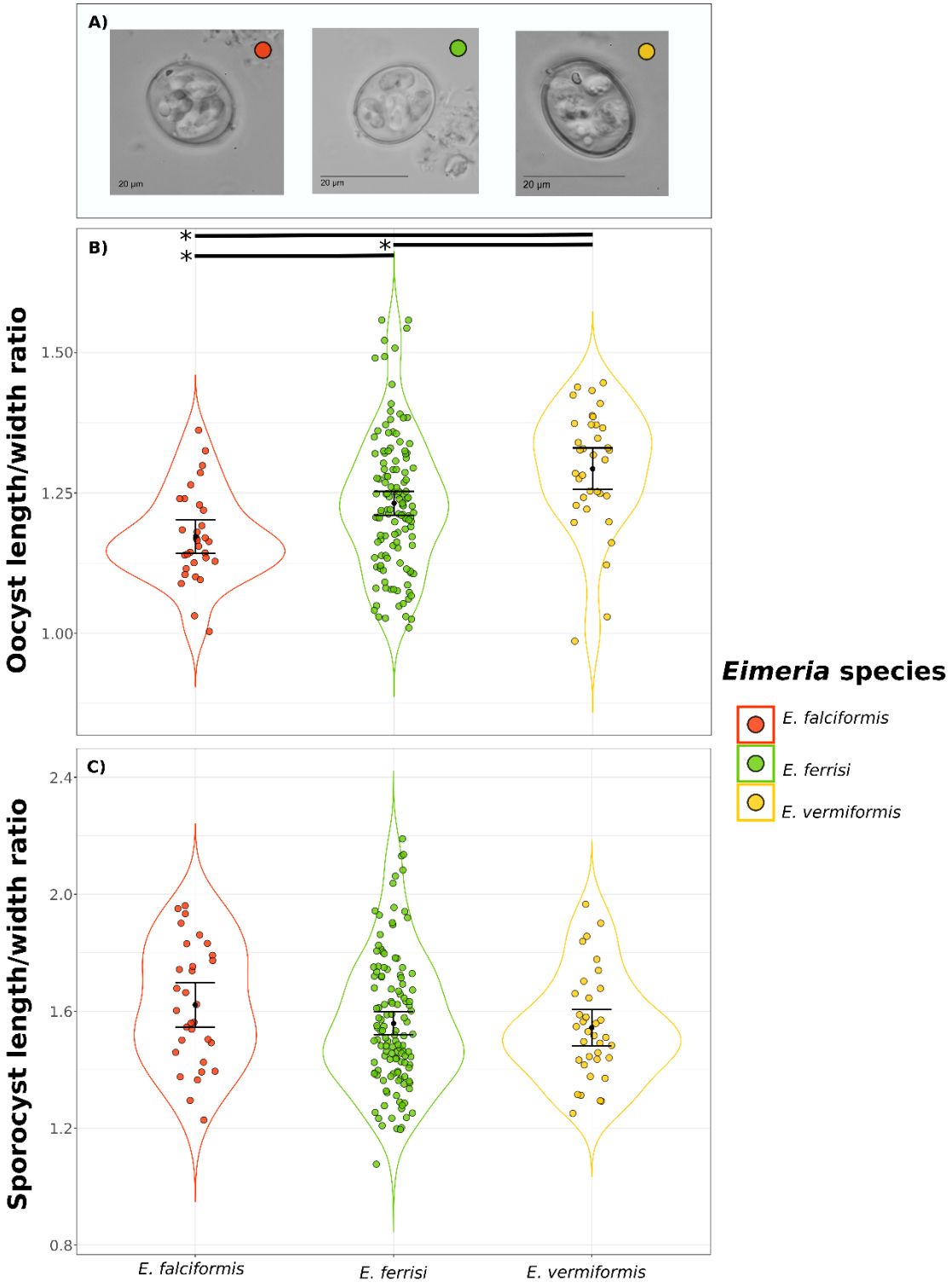



**Figure 2.3 Morphological and morphometrical characteristics of *Eimeria* oocyst isolated from *Mus musculus*.** a) Photomicrographs at 1000x amplification of *Eimeria* oocyst from the three species isolated from *Mus musculus* (red = *E. falciformis*; green = *E. ferrisi* and yellow = *E. vermiformis*). Length/Width ratio from b) oocyst and c) sporocysts corresponding to each species (*E. falciformis n*= 31; *E. ferrisi n*= 127 and *E. vermiformis n*= 35). Mean +/- 95% Confidence Interval is plotted. * Represent significant differences (Tukey HSD, $p<0.05$).

### 2.4.4 Proximal-distal occurrence of infection and double infections

We detected DNA from endogenous stages by qPCR in 27 of 163 samples analysed (Supplementary data S2.6). We differentiate detection between small and large intestine, analysing ileum as the most distal tissue of the small intestine and cecum as the most proximal tissue of the large intestine. Detection was either limited just to cecum (*n* = 19), to ileum (*n* = 2) or possible in both tissues (*n* = 6). Infections in cecum were identified as *E. falciformis* (*n* = 4), *E. ferrisi* (*n* = 17) and *E. vermiformis* (*n* = 1) Detections in ileum (*n* = 2) were identified as *E. vermiformis*. In two mice positive in both tissues, it was possible to identify *E. ferrisi* in cecum and *E. vermiformis* in ileum, providing evidence that these animals presented a double infection (that is, simultaneous infections with different isolates; Figure 2.4).

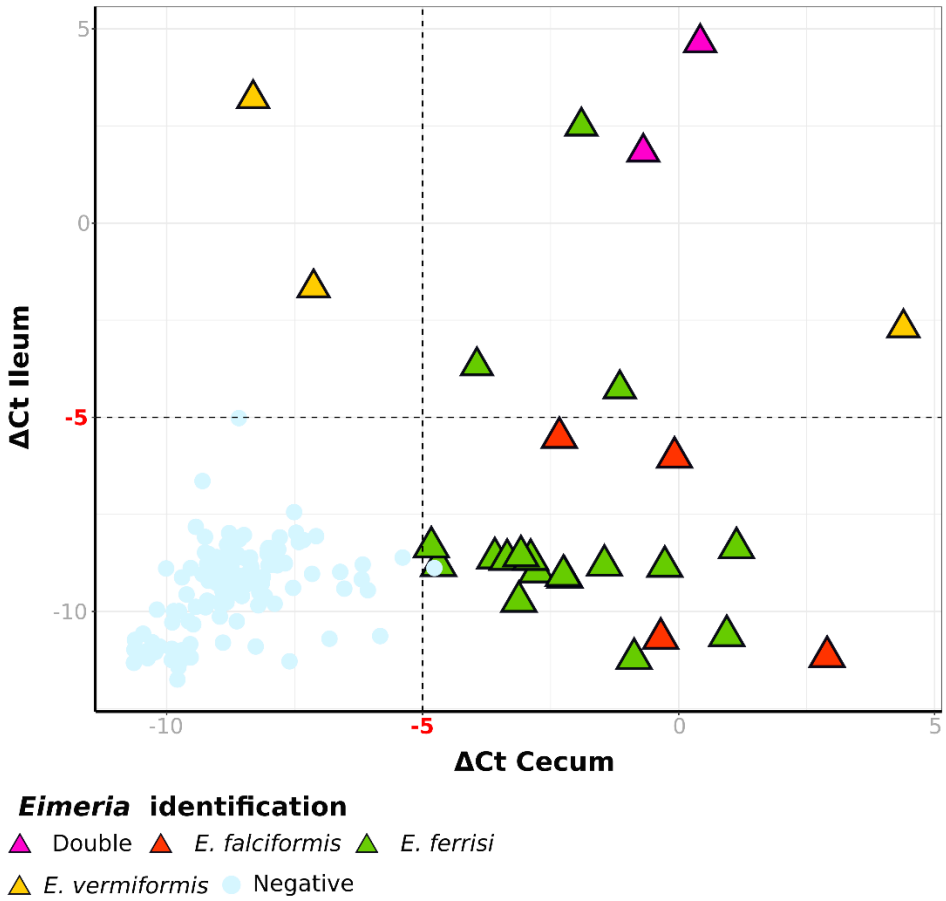



**Figure 2.4 qPCR detection of intracellular stages of *Eimeria* in cecum and ileum from *Mus musculus*.** -Delta Ct value ($Ct_{Mouse}$ - $Ct_{Eimeria}$) from each tissue for 164 mice are plotted on the graph. The dotted lines indicate the threshold of -5, values above the line are considered positive for the corresponding tissue. Circles represent negative samples, triangles indicate samples with *Eimeria* species identification and colors correspond to the *Eimeria* species identified in those samples.

### 2.4.5 Amplification efficiency of different markers

As *Eimeria* detection by PCR and the determination of species identity could be biased especially in cases of double infections, we analysed differences in amplification efficiency of the three primer pairs used for the molecular identification of *Eimeria* species (Table 2.2). In a cross-validation approach we compare the likelihood to amplify a maker given the species identification with the other marker. The amplification and sequencing efficiency of the Cocci_COI primer pair was significantly higher for *E. ferrisi* isolates (logistic regression, $p < 0.001$) than for *E. falciformis* isolates (the letter determined by the 18S marker). Using the novel primer pair Eim_COI_M_F/R the sequencing results were complemented and we detected no significant differences in the combined amplification efficiency for both primers (logistic regression; $p = 0.62$).

Similarly, we did not detect significant differences in the probability to obtain an 18S sequence (using the 18S_EF/R primer pair) between *Eimeria* species as determined by (combined) COI assessment (logistic regression; $p = 0.25$). Differences in PCR efficiency for Cocci_COI make it likely that markers amplify different species in case of double infections in a single isolate. Both detection by flotation and diagnostic (Ap5) PCR significantly increase the likelihood amplification for all primers (COI or 18S). The statistical models for biased amplification hence also confirm that both detection methods provide complementary results, even while controlling for detection with the other method (Table 2.2).

**Table 2.2. Statistical models used to analyse primer preference to different *Eimeria* species**

| Predictors | Model 1 (Cocci_COI) | Model 2 (Cocci_COI + Eim_COI) | Model 3 (18S_EF/R) |
|---|---|---|---|
| Intercept | -5.80 ***<br>(1.11) | -1.55<br>(1.27) | -1.40<br>(0.87) |
| Flotation | 1.64 **<br>(0.52) | 2.15 ***<br>(0.54) | 1.59 **<br>(0.51) |
| Identification *E. ferrisi* (other marker) | 5.10 ***<br>(1.14) | 0.28<br>(1.35) | 0.32<br>(0.63) |
| Identification *E. vermiformis* (other marker) | 17.23<br>(1385.38) | 11.98<br>(1385.38) | 1.71<br>(1.33) |
| No identification (other marker) | 0.97<br>(0.92) | -3.57 **<br>(1.23) | -3.44 ***<br>(0.87) |
| Identification *Eimeria* spp. (COI) | | | -14.80<br>(1455.40) |
| *n*<br>AIC<br>BIC<br>Pseudo $R^2$ | 378<br>128.09<br>151.70<br>0.77 | 378<br>119.18<br>142.79<br>0.82 | 378<br>143.39<br>170.93<br>0.70 |

The upper number represents the estimate and numbers in brackets represent standard error for each predictor.
Intercept is Eimeria falciformis identification with other marker
Other marker refers to 18S based identification for COI models or vice versa
*n* (Total number of samples), AIC (Akaike information criterion), BIC (Bayesian information criterion).
*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

### 2.4.6 Prevalence of the different *Eimeria* species

Genotyping amplification and sequencing with either 18S or COI primers (or both) was possible for samples in which infections had been detected (*n*= 82) (Figure 2.1B). Both COI and 18S genotyping PCRs hence support detection by flotation and diagnostic PCRs in this subset of samples. Furthermore amplification of both or either one maker was fully sufficient to assign the isolate to an *Eimeria* species (Figure 2.2), allowing us to resolve prevalence on the species level.

These corrections and controls allow us to determine prevalence at the species level: *E. ferrisi*

was identified at a higher prevalence of 16.7% (63/378, 95%CI = 13.2 — 20.7) in comparison to *E. falciformis* (16/378, 4.2% [95%CI = 2.6—6.8]) and *E. vermiformis* (7/378, 1.9% [0.9—3.8]).

Considering prevalence at the level of farms, *E. ferrisi* was detected in 29.2% (28/96, 95%CI = 20.7 – 39.0), *E. falciformis* in 12.5% (12/96, 95%CI = 7.1 – 20.7) and *E. vermiformis* in 7.3% (7/96, 95%CI = 3.5 – 14.4) of sampled localities. 25 (of in total 96) farms, had mice with single *Eimeria* species detected, and 10 had more than one species detected. In all cases *E. ferrisi* was detected (5 farms with *E. ferrisi* – E. *falciformis*, 3 farms with *E. ferrisi* – *E. vermiformis* combination, and 2 farms with the three species). Mice presenting double infections were caught at farms at which infections with the both *Eimeria* species were found in other mice independently.

We used the number of mice caught per farm as a proxy for population density, assuming roughly equal trapping effort at all localities. We then question whether population density affects prevalence by testing differences in the likelihood of a mouse individual to be infected dependent on that population density. We detect that the likelihood of infection is significantly increased for both *E. ferrisi* and for *E. falciformis* (logistic regression, $p < 0.05$; Table 2.3). Infection with *E. falciformis* got more likely by 19%, infection with *E. ferrisi* by 14% with each mouse caught at the same locality. We also included the detection of other *Eimeria* species in the model for each species and did not find a significant influence ($p > 0.05$) on likelihood of infection.

**Table 2.3. Statistical models used to analyse factors influencing infection to different *Eimeria* species**

| Predictors | Model 1 (*E. ferris* infection) | Model 2 (*E. falciformis* infection) | Model 3 (*E. vermiformis* infection) |
|---|---|---|---|
| Intercept | -1.62 ***<br>(0.35) | -3.32 ***<br>(0.61) | -3.99 ***<br>(0.83) |
| Total caught mice | 0.13 *<br>(0.06) | 0.18 *<br>(0.08) | 0.11<br>(0.10) |
| *E. ferris* infection | - | 0.92<br>(0.70) | 0.29<br>(1.06) |
| *E. falciformis* infection | 0.92<br>(0.69) | - | 1.60<br>(0.92) |
| *E. vermiformis* infection | 1.61<br>(0.91) | 0.41<br>(1.03) | - |
| *n* | 104 | 104 | 104 |
| AIC | 120.99 | 71.61 | 52.03 |
| BIC | 131.57 | 82.18 | 62.61 |
| Pseudo $R^2$ | 0.18 | 0.19 | 0.17 |

The upper number represents the estimate and numbers in brackets represent standard error for each predictor.
*n* (Total number of samples), AIC (Akaike information criterion), BIC (Bayesian information criterion).

## 2.5 Discussion

In this study we identify *Eimeria* species in wild commensal populations of house mice (*Mus musculus*). We show that detection and identification of this group of rodent coccidia can be challenging and propose to complement classical coprological assessment with molecular tools: a highly sensitive detection PCR, genotyping PCRs for species identification and qPCR for localization and detection of double infections. Based on this we identified three different *Eimeria* species in the house mouse: *E. ferrisi*, *E. falciformis* and *E. vermiformis.* Morphological characteristics and preferential occurrence were congruent with the assignment of isolates to the above species. We use our results to show a positive effect on host density on prevalence of *E. ferrisi* and *E. falciformis*.

Few studies report prevalence of *Eimeria* in wild populations of *Mus musculus*. Prevalences range from 3% to 40% for isolates classified either as *E. falciformis*, *E. ferrisi* or *E. vermiformis*[112,171,178,179,197]. Other studies make no assessment at the species level (detection

as *Eimeria* spp)[167,181,182,198].

A recent study in rodents (other than *Mus musculus*) in central Europe reported an *Eimeria* spp. prevalence of 32.7% based on coprological observations[91]. At 37.6% the overall prevalence for all *Eimeria* species in our study in house could be considered high in comparison to all these studies.

While flotation is the most commonly used for detection and quantification of coccidia[199–201], we here used a complementary approach of flotations and diagnostic PCR. We observed relatively large discrepancies between both methods. Flotation and counting of oocysts has a relatively high limit of detection[202,203], explaining negative findings in oocyst flotations positive for PCR. Negative PCR results for samples with visible oocyst in flotations could be a result of a failure to break oocyst walls during DNA extractions and/or faecal PCR inhibition[204] (Raj et al., 2013). Importantly, tested but could not find any species-specific bias in both methods making e.g. relative species prevalences reliable**.**

Traditional identification of *Eimeria*, depends on the expertise to recognise the morphology of sporulated oocyst[112]. We show that interpretation of morphometrical data is complex due overlap between species while measurement means to agree with literature (Table 2.1). Considering the challenges of identification and characterisation of *Eimeria* isolates from field samples, we conclude that characterisation of *Eimeria* species requires molecular markers and phylogenetic analysis.

Sequence identity of our isolates to reference sequences from *Eimeria* species previously described in *M. musculus* was above 98% for COI, which is sometimes assumed to have sufficient corresponding differences within species of *Eimeria*[205]. We confirm taxonomic assignment based on highly supported maximum likelihood and Bayesian phylogenetic clustering of 18S and COI sequences. Moreover, the three identities of the three *Eimeria* species were supported by phenotypic characteristics: morphometry of oocysts and tissue occurrence of the infection.

For some *Eimeria* species precise tissue localization is described based on histology or electron microscopy[176,206]. Both methods provide detailed information on developmental stages, but are also time consuming and require a high level of expertise. As an alternative to determine (only) the rough occurrence of the infection along the proximal-distal axis of the intestine, a DNA based qPCR method allowed us not only to detect the presence of *Eimeria*, but also to estimate tissue specific intensity of infection. The qPCR targets a single-copy nuclear gene from the host and a mitochondrial gene of the parasite present in multiple copies

(up to 180[143]) to increase sensitivity for *Eimeria* detection.

While infection with rodent *Eimeria* in general can be limited to the duodenum and jejunum[149], house mouse *Eimeria* have been described to be mostly found in either the small or the large intestine[112,168,171,176]. Using ileum, as the most distal part of the small intestine, and cecum, as the most proximal of the large intestine, we aimed to provide the most stringent test for differences in the site of infection possible: strong infections can be expected to spread in the neighboring tissue, but one could still expect the primary tissue to be more strongly infected. Additionally, genotyping DNA derived from these tissues allowed us to detect double infections with *E. vermiformis* in the small intestine and *E. ferrisi* in the large intestine. Localization generally agrees with previous descriptions for the isolates we identified as *E. ferrisi*, *E. falciformis* and *E. vermiformis* by phylogenetic clustering. Co-infections have been reported previously in *A. sylvaticus*[207] or in large populations of grey and red squirrels[208]. To our knowledge we provide the first report of double infections in wild populations of *Mus musculus*.

Double infections can be problematic for identification of species by genotyping. Simultaneous infection of the caecum with *E. ferrisi* and *E. falciformis* would not be recognized with our qPCR method. We showed that amplification of COI with the commonly used primer pair Cocci_COI[72] is differentially efficient for different *Eimeria* isolates. This primer preference can lead to a misidentification in double infections due to the generation of “chimeric” isolates that present different and contradictory information for different markers. Attention to such discrepancies is needed when collating database sequences and when developing multi-marker approaches in general. For rodent *Eimeria* systems, we develop an alternative COI primer pair and find no evidence for differential amplification bias in our cross-validation of the different primer pairs.

*E. ferrisi* is by far the most prevalent species in our study area infecting *M. musculus*. Concerning the house mouse hybrid zone we find infections in *M. m. domesticus, M. m. musculus* and hybrids, suggesting that there are no strict geographical or host subspecies constraints for this species. Population structure for *E. ferrisi* (which could in turn correspond to host subspecies[209,210] cannot be found at the resolution the analysed markers provide.

We found that prevalences of *E. ferrisi* and *E. falciformis* increase with increasing host density at the level of farms. This is in agreement with predictions from epidemiology that in large and dense populations contact rates increase[211] and microorganisms with direct transmission become more prevalent. Such prevalence – host density relationships have been well documented for Hantavirus infections in Bank vole (*My. glareolus*)[212–214]. In free-living populations of house mice increased host density has been observed to result in higher

prevalence of Murine Cytomegalovirus (MCMV)[215]. For eukaryotic parasites the prevalence of cestodes and nematodes has been described to be host density dependent in wild and laboratory rodents[216,217]. We here document such host density – prevalence relationship for the first time at a species level in *Eimeria* of house mice.

We suggest that species level identification of parasites in wildlife systems will help to assess such questions in more detail and is absolutely required for other questions. For example virulence-prevalence trade-off[218,219] can only be assessed at the species level. In our system one would predict a lower virulence for the prevalent *E. ferrisi* compared to *E. falciformis* and *E. vermiformis.* We have indications from laboratory experiments that such a lower virulence of *E. ferrisi* might be observed compared to *E. falciformis*[144], while contrary results have been reported before[220]. We consider this an observation warranting further research.

In conclusion we argue that *Eimeria* in wildlife populations should be identified more frequently at the level of species previously described by taxonomists. We propose to integrate a set of simple methods into a reproducible procedure to achieve this aim. For Coccidians, as important parasites of vertebrates, only species specific assessment will allow to test hypotheses in evolution, ecology and epidemiology.

# Chapter 3. Generalist *Eimeria* species in rodents: Multi-locus analysis indicate inadequate resolution of established markers

**Published article**

Jarquín-Díaz, V. H., Balard, A., Mácová, A., Jost, J., Roth von Szepesbéla, T., Berktold, K., Tank, S., Kvičerová, J., & Heitlinger, E.

*Ecology and evolution* (2020), *10*(3), 1378–1389.

https://doi.org/10.1002/ece3.5992



**Author contributions:**
VHJD and EH designed the project and obtained funding. VHJD, AM, TRS, JJ, KB, ST and JK obtained data, VHJD, AB and EH designed the analysis, VHJD, AB and EH performed the analysis. VHJD, EH and JK interpreted the results. VHJD and EH wrote the manuscript with contributions from all other authors. EH supervised the project.

## 3.1 Abstract

Intracellular parasites of the genus *Eimeria* are described as tissue/host specific. Phylogenetic classification of rodent *Eimeria* suggested that some species have a broader host range than previously assumed. We explore whether *Eimeria* spp. infecting house mice are misclassified by the most widely used molecular markers due to a lack of resolution, or whether, instead, these parasite species are indeed infecting multiple host species.

With the commonly used markers (18S/COI), we recovered monophyletic clades of *E. falciformis* and *E. vermiformis* from *Mus* that included *E. apionodes* identified in other rodent host species (*Apodemus* spp.*, Myodes glareolus,* and *Microtus arvalis*). A lack of internal resolution in these clades could suggest the existence of a species complex with a wide host range infecting murid and cricetid rodents. We question, however, the power of COI and 18S markers to provide adequate resolution for assessing host specificity. In addition to the rarely used marker ORF470 from the apicoplast genome, we present multi-locus genotyping as an alternative approach. Phylogenetic analysis of 35 nuclear markers differentiated *E. falciformis* from house mice from isolates from *Apodemus* hosts. Isolates of *E. vermiformis* from *Mus* are still found in clusters interspersed with non-*Mus* isolates, even with this high resolution data.

In conclusion, we show that species-level resolution should not be assumed for COI and 18S markers in Coccidia. Host-parasite co-speciation at shallow phylogenetic nodes, as well as contemporary coccidian host ranges more generally, are still open questions that need to be addressed using novel genetic markers with higher resolution.

**Keywords**

Rodents, *Eimeria*, phylogenetics, COI, 18S, multi-locus sequence typing

## 3.2 Introduction

Coccidians of the genus *Eimeria* have been described as monoxenous, intracellular parasites[108,124,221]. Two different characteristics extensively used to delineate *Eimeria* species are their assumed high degree of host and tissue specificity. It is not clear, however, whether host specificity is the same for *Eimeria* species infecting hosts in different clades. *Eimeria* species of rodents show a degree of host specificity[117,146,178,222] but individual isolates can experimentally infect different species and even genera of rodents[147,223].

Descriptions of *Eimeria* species are based on the size and shape of sporulated oocysts and their internal structures. The life cycles of a few species has additionally been studied and data

on their dynamics (e.g. the patent period, the time before oocysts are shed in faeces) are available[112,136,175,224–229]. For field studies, the morphology of sporulated oocysts alone is considered insufficient to infer species identity because of inadequate reference descriptions[89,230].

Genetic markers from nuclear (nu) and mitochondrial (mt) genomes, and less frequently of the apicoplast (ap) genome, have been used to complement morphological taxonomy with phylogenetic analyses[77,130,137,159,162]. Based on the assumption of host specificity of individual *Eimeria* species, phylogenetic analysis of nuclear small subunit ribosomal (18S) rRNA and cytochrome *c* oxidase I (COI) fragments supports predominant host-parasite co-speciation[78]. Species infecting rodents, however, are found in two separate clades, generating marked discrepancy between parasite and host phylogeny at deeper nodes[90]. At shallow nodes of the phylogeny for rodent coccidians, cases of host generalism have been suggested[91]. Host specificity of *Eimeria* species infecting rodents is not as undisputed as in other hosts such as poultry[231] or rabbits[132]. Kvičerová and Hypša[90] suggested that adaptation rather than co-speciation is shaping rodent-*Eimeria* co-phylogenies. Mácová *et al.*[91], added that host ecology and distribution may favour host-switches among closely related rodent species. A high specificity of *E. apionodes* naturally infecting *Apodemus flavicollis* was originally suggested based on failed attempts to experimentally infect other rodents: *Myodes (Clethrionomys) glareolus*, *Microtus arvalis*, or *Mus musculus*[232]. It is, however, unclear if this result holds for the multiple isolates that have been assigned as *E. apionodes*.

We studied wild populations of *Mus musculus* and other rodents to assess the diversity of *Eimeria* isolates at shallow depth of phylogenetic relationships. We test host specificity based on phylogenetic analysis using established markers (nu 18S, mt COI and ap ORF470). We question how far these markers are polymorphic enough to resolve between genetic clusters with different host usage (and whether a negative result for genetic differentiation therefore suggests (cases of generalism). We develop and apply multi-locus sequence typing to disentangle relationships unresolved by 18S and COI markers.

## 3.3 Material and methods

### 3.3.1 Origin of samples

DNA was extracted from the colon content or gastrointestinal tissue of house mice (*Mus musculus*) infected with *Eimeria*. These samples came from rodents captured in farms and private properties in the German federal states of Mecklenburg-Vorpommern, Bavaria and Brandenburg (capture permit No. 2347/35/2014) and in Bohemia (Czech Republic) between

2014 and 2017[233]. Additionally, DNA from gastrointestinal tract, tissue or faeces of *Apodemus* spp. from different regions in Europe (including areas overlapping with those sampled for house mice) were also included[91] (Figure 3.1) (Supplementary data S3.1).


**Figure 3.1 Location of rodent samples.** *Mus musculus* samples were collected from the German federal states of Mecklenburg-Vorpommern, Bavaria and Brandenburg and in Bohemia (Czech Republic). Non-*Mus* samples were collected from different countries within Europe. Colour in the points indicate the host species.

### 3.3.2 Host identification

Rodents were first identified visually based on their morphology. Identification of *Mus musculus* at the sub-species level was confirmed based on a set of previously described markers[184]. In order to confirm the species of non-Mus rodents, a fragment of cytochrome b (~900 bp) was

amplified from host DNA. PCRs were performed according to the protocols described by Reutter *et al.*[234] for *Apodemus* spp., Abramson *et al*.[235] (primers UCBO_F/LM_R) and Jaarola and Searle[236] (primers L14641M/H15408M) for rodents belonging to the subfamily Arvicolinae (*Myodes* spp. and *Microtus* spp.).

### 3.3.3 PCR amplification (nu SSU 18S rRNA, mt COI and ap ORF470)

For phylogenetic analysis, nuclear small subunit ribosomal DNA (18S; ~1,500 bp), a fragment of the mitochondrial cytochrome *c* oxidase subunit I (COI; ~800 bp) gene and apicoplast ORF470 (~800 bp) were amplified using primers previously reported by Kvičerová *et al.*[132], Ogedengbe et al. [72] and Zhao and Duszynski[131], respectively.

When COI failed to amplify with this protocol, an alternative pair of primers was used: Eim_COI_M_F (ATGTCACTNTCTCCAACCTCAGT) and Eim_COI_M_R (GAGCAACATCAANAGCAGTGT). These primers amplify a ~700 bp fragment of COI and were designed based on the mitochondrial genome of *E. falciformis* (CM008276.1)[143,233].

PCR reactions were carried out in a Labcycler (SensoQuest GmbH, Göttingen, Germany) using 0.025 U/µL of DreamTaqTM DNA Polymerase (Thermo Scientific, Waltham, USA), 1X DreamTaq Buffer, 0.5 mM dNTP Mix, 0.25 µM from each primer and 1 – 20 ng/µL of DNA template in 25 µL reaction. A concentration of 0.25 mM dNTP Mix and a supplementation with 0.5 mM MgCl2 was used for the ap ORF470 amplification. The thermocycling protocol consisted of 95 °C initial denaturation (4 min) followed by 35 cycles of 92 °C denaturation (45 s), annealing at 52°C (30 s/Eim_COI); 53 °C (45 s/18S); 55 °C (30 s/COI); 50 °C (45 s/ORF470); 72 °C extension 90 s (18S/ORF470), 20 s (COI/Eim_COI), and a final extension at 72 °C (10 min). DNA from oocysts of *E. falciformis* BayerHaberkorn1970 and DNA from colon content of a non-infected lab (NMRI) mouse were used as positive and negative controls, respectively.

All PCR products from nu 18S, mt COI and ap ORF470 of the expected size were purified using the SAP-Exo Kit (Jena Bioscience GmbH, Jena, Germany), and sequenced in both directions by LGC Genomics (Berlin, Germany). Quality assessment and sequence assembly was performed in Geneious v6.1.8. All sequences were submitted to the NCBI GenBank database (Accession numbers: nu SSU 18S rRNA [MH751925-MH752036, MK246860-MK246868 and MK625202-MK625210]; mt COI [MH777467-MH777593, MH755302-MH755324, MK257106-MK257114 and MK631866-MK631868] and ap ORF470 [MH755325-MH755450, MK257115-MK257125 and MK631869-MK631884]).

### 3.3.4 Phylogenetic analysis and inference of intraspecific genetic diversity

Datasets for each gene and a concatenated alignment (nu 18S, mt COI and ap ORF470) were created adding closely related reference sequences available in the GenBank (supplementary data S3.2).

Protein coding sequences (mt COI and ap ORF470) were aligned by translation using the Multiple Align algorithm and translation frame 1 with the genetic code for “mold protozoan mitochondrial”, 18S sequences were aligned using MUSCLE[187], both through Geneious v6.1.8.

Phylogenetic trees for all datasets were constructed using Maximum Likelihood (ML) and Bayesian inference (BI) methods, implemented in PhyML v3.0[188] and MrBayes v3.2.6[189,190], respectively. Sequence evolution models most appropriate for each dataset were determined in JModelTest v2.1.10[191]. For ML trees, a bootstrap analysis with 1,000 replicates was performed, whereas MCMC for BI was run with two cold and two hot chains for 1,000,000 generations or until the split freq value was below 0.05. The concatenated dataset was analysed using partitions and locus-specific models. Trees were visualized with FigTree v1.4.2[192]. A haplotype network of mt COI sequences was inferred using a codon-based alignment trimmed to 500 bp available for all isolates. Haplotypes frequencies were calculated and a network was constructed with the R package “pegas” v0.11[237].

### 3.3.5 Multi-marker genotyping PCR and high throughput sequencing

Samples positive for *E. falciformis* and *E. vermiformis* from *Mus musculus* and *Eimeria* spp. from Apodemus with indistinguishable 18S and COI sequences were used for a multi-marker amplification using the microfluidics PCR system Fluidigm Access Array 48 x 48 (Fluidigm, San Francisco, California, USA). We used target specific primers (Supplementary data S3.3) that were designed based on the genome of E. *falciformis*[143] to amplify exons of nuclear genes (Supplementary data S3.4) and coding and non-coding regions from the apicoplast genome (Supplementary data S3.5). Library preparation was performed according to the protocol Access Array Barcode Library for Illumina Sequencers (single direction indexing) as described by the manufacturer (Fluidigm, San Francisco, California, USA). The library was purified using Agencourt AMPure XP Reagent beads (Beckman Coulter Life Sciences, Krefeld, Germany). Quality and integrity of the library was confirmed using the Agilent 2200 Tapestation with D1000 ScreenTapes (Agilent Technologies, Santa Clara, California, USA). Sequences were generated at the Berlin Center for Genomics in Biodiversity Research (BeGenDiv) on the Illumina MiSeq platform (Illumina, San Diego, California, USA) in two runs, one using “v3 chemistry” with 600 cycles, the other “v2 chemistry” with 500 cycles. All sequencing raw data

can be accessed through the BioProject PRJNA548431 in the NCBI Short Read Archive (SRA).

### 3.3.6 Bioinformatic analysis of multi-locus sequence typing

Screening and trimming of sequencing reads was performed using the package dada2 v1.2.1[238]. All reads were trimmed to 245 bases, while allowing a maximum of 4 expected errors (maxEE). Sorting and assignment to amplicons was performed with the package MultiAmplicon v0.1[239] and the most abundant sequence was recorded for each marker in each sample (recording but disregarding minority sequence in non-clonal infection for further analysis; see Supplementary data S3.6). Sequences were aligned using the function "AlignSeqs" from the package DECIPHER v2.10.0[240] and non-target sequences were excluded from alignments if >20% divergence was observed with other sequences (such as in cases off-target amplification of mostly bacterial sequences). Alignments were controlled for the absence of insertions/deletions (indels) that distort the open reading frame. Prevalent multiple-of-3-mere indels corresponding to homopolymeric amino acid repeats (HAARs[143]) of diverse length were coded as missing data due to their unclear model of evolution. The function "dudi.pcr" from the packages ade4 v1.7-13[241] and adegenet v2.1.1[242] was used to visualize genetic distances between samples based on all markers. The code for this pipeline is available at https://github.com/VictorHJD/AA_Eimeria_Genotyping.

The alignments of the concatenated sequences were then exported. The number of informative sites was summarized using the tool DIVEIN[243] and phylogenetic trees were computed by Bayesian inference in MrBayes v3.2.6[189,190]. A partitioned model was implemented to estimate the tree considering each gene separately. The analysis was performed with two runs, with 1,000,000 generations leading to a split frequency value below 0.05, and 200,000 generations were discarded as burn-in when estimating posterior probability. Additionally, Maximum Likelihood trees were inferred with 1,000 bootstrap replicates in PhyML v3.0[188].

The topology of ML and BI trees was compared and summarized into a consensus tree with minimum clade frequency threshold of 0.95 using the program SumTrees v4.3.0[244].

## 3.4 Results

### 3.4.1 Established markers don't recover clades corresponding to species with different host-usage

We performed phylogenetic analyses using nuclear, mitochondrial, and apicoplast markers to assess the clustering of our sequences into groups of previously described species.

We inferred a phylogenetic tree of nu 18S based on 215 sequences (509 – 1,795 bp). Of these, 111 from parasites in house mice (*M. musculus*) (3 from ileum tissue, 16 from cecum tissue and 92 from colon content) and 18 from parasites in non-*Mus* rodents were generated in the present study (3 from ileum tissue, 3 from cecum tissue, 3 from colon content, and 9 from feces). To test for host specificity of house mouse *Eimeria* we included reference sequences from related *Eimeria* species described in murid and cricetid rodents. *Isospora* sp. sequences identified in *Talpa europaea* moles were used as an outgroup. Both ML and BI rooted trees shared the general topology (Figure 3.2).

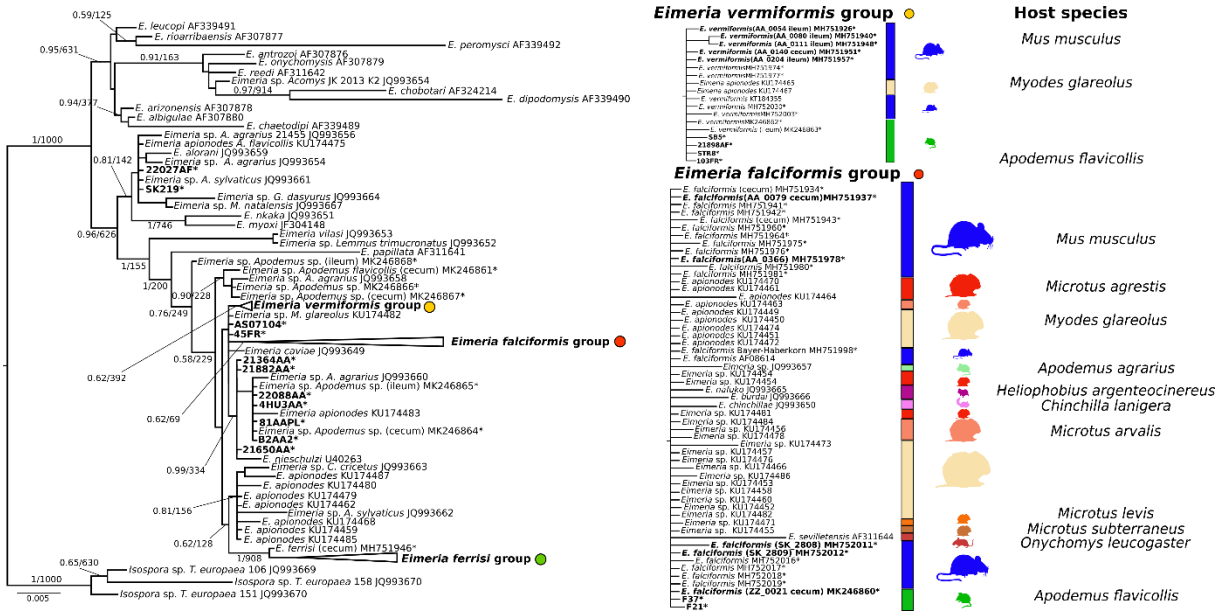



**Figure 3.2 Phylogenetic trees inferred from nuclear small ribosomal subunit (SSU 18S rRNA).** Phylogenetic tree based on SSU 18S rRNA sequences. Numbers in the branches represent Bayesian posterior probability and bootstrap value. The three collapsed groups cluster *Eimeria* sequences from *M. musculus* of this study. Reference sequences from other rodents were included. The scale bar represents sequence divergence. Hosts for closely related sequences of *E. falciformis* and *E. vermiformis* are indicated in the expanded form of the group. * Represent sequences generated in the present study. Tissue of origin is indicated in brackets. Sequences in bold were included in the multi-marker phylogenetic inference.

The sequences derived from *Mus musculus* samples clustered in three well supported monophyletic groups: one comprising reference sequences of *E. falciformis* (*E. falciformis* group), another of *E. ferrisi* (*E. ferrisi* group), and the third of *E. vermiformis* (*E. vermiformis* group). All three groups, however, included sequences of *Eimeria* from other cricetid and murid hosts without showing internal sub-structure linked to the observed host species infected (host-usage) (Figure 3.2).

The phylogenetic tree of mt COI was based on 233 sequences (381 – 804 bp), 149 of which were obtained from *Eimeria* infecting house mice (3 from ileum, 16 from cecum tissue and 130 from colon content) and 12 from non- *Mus* rodents in our study (2 from ileum, 1 from cecum, 6 from colon content, and 3 from feces) (Figure 3.3). Similar to 18S, COI sequences derived from house mice clustered in three monophyletic groups including reference sequences of *E. falciformis* (*n*= 26), *E. ferrisi* (*n*= 109) and *E. vermiformis* (*n*= 13). Groups of *E. falciformis* and *E. vermiformis* also include sequences derived from *Eimeria* isolates of common voles (*Mi. arvalis*), bank voles (*My. glareolus*), short-tailed voles (*Mi. agrestis*), yellow necked mice (*A. flavicollis*) or wood mice (*A. sylvaticus*). In addition to our isolates from *M. musculus*, the *E. ferrisi* group contain sequences of *E. burdai* and *E. nafuko,* species described from sub Saharan mole rats (*Heliophobius argenteocinereus*). Again, the clades do not show further sub-structure indicative of host usage (Figure 3.3).

A phylogenetic tree of ORF470 was based on 172 sequences (Figure 3.4) and showed a similar topology to the COI and 18S trees. Sequences derived from *Eimeria* isolates from *Mus musculus* (*n*= 125) also clustered into the same three groups. For this marker, the number of sequences available in databases from other cricetid and murid rodents is very limited, and none of the available sequences clustered within the highly supported “species clusters” of our isolates. In contrast to nu 18S and mt COI, our newly generated sequences from isolates detected in *A. flavicollis* and *A. sylvaticus* formed separate clusters that were basal to the *E. falciformis* group (*n*= 3), and outside of the *E. vermiformis* group (*n*= 4) (Figure 3.4).

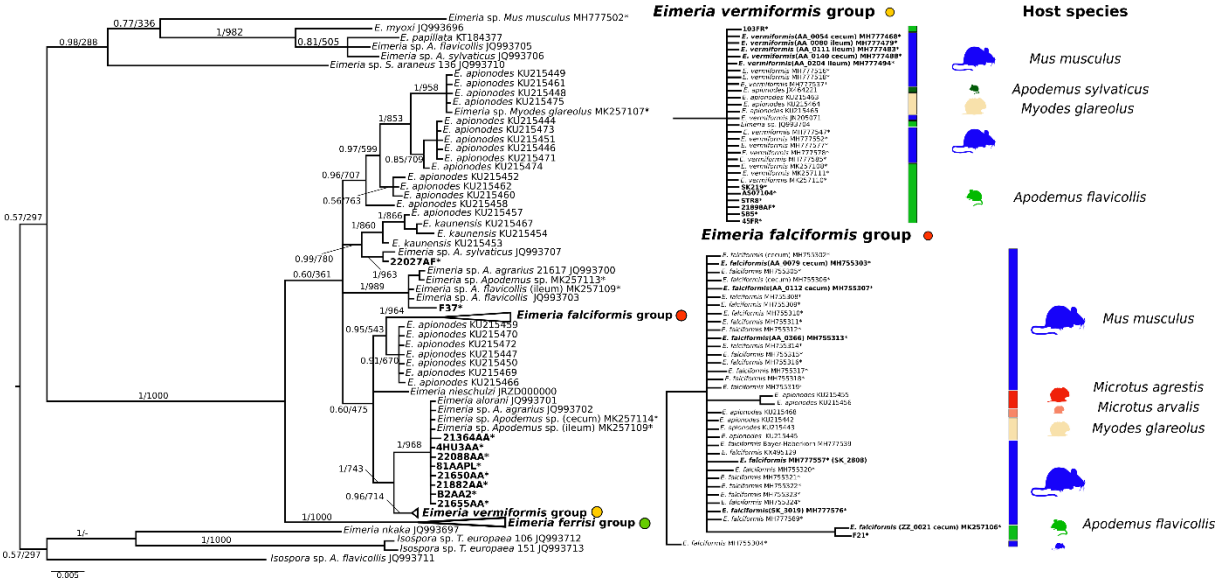



**Figure 3.3 Phylogenetic trees inferred from mitochondrial cytochrome c oxidase (COI) sequences.** Phylogenetic tree based on COI. Numbers in the branches represent Bayesian

posterior probability and bootstrap value. The three collapsed groups cluster *Eimeria* sequences from *M. musculus* of this study. Reference sequences from other rodents were included. The scale bar represents sequence divergence. Hosts for closely related sequences of *E. falciformis* and *E. vermiformis* are indicated in the expanded form of the group. * Represent sequences generated in the present study. Tissue of origin is indicated in brackets. Sequences in bold were included in the multi-marker phylogenetic inference.

To combine all available information into a single phylogenetic analysis we used a concatenated alignment. In the tree constructed from this alignment (Supplementary data S3.7), the clusters from *E. vermiformis* and *E. ferrisi* observed in the individual phylogenies with 18S, COI and ORF470 were confirmed in the concatenated tree. Sequences from the *E. falciformis* group were found in an unresolved basal position with *E. apionodes* isolates derived from *Myodes* sp. and *Microtus* sp.. This result probably indicates conflicting signals for different markers and missing data.

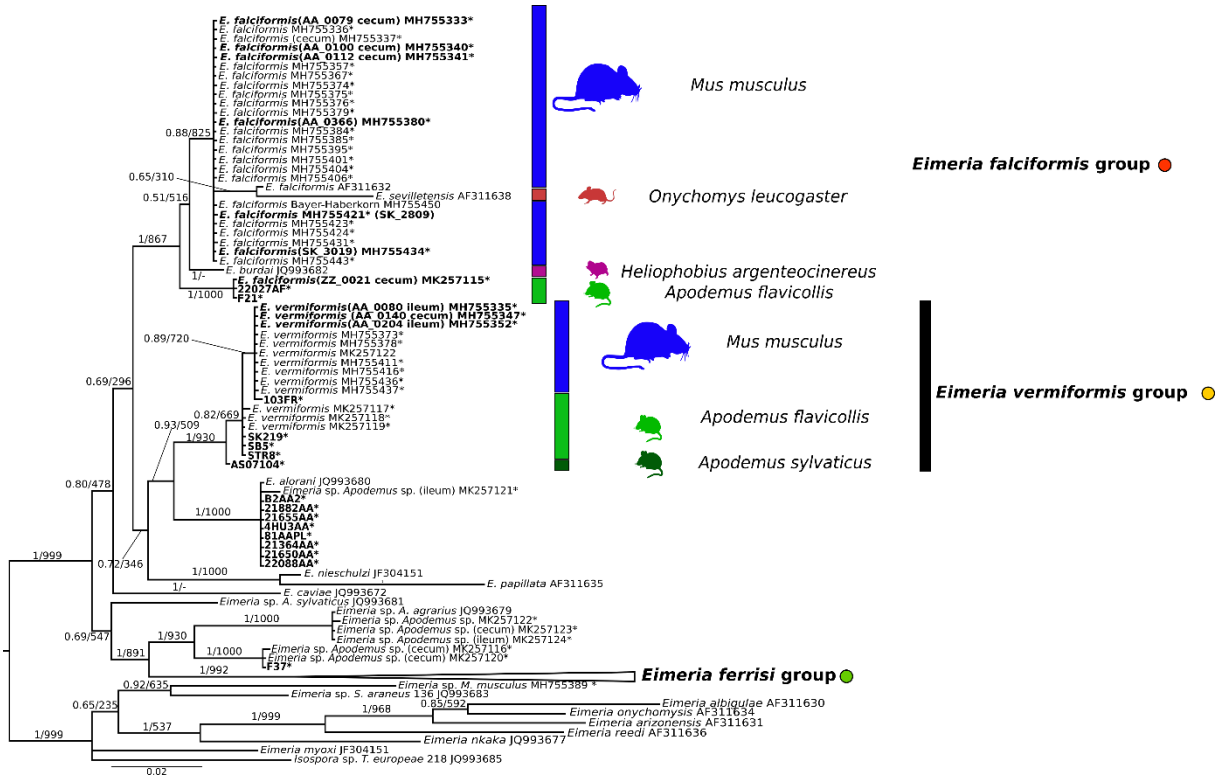



**Figure 3.4 Phylogenetic trees inferred from apicoplast open reading frame 470 (ORF470) sequences.** Phylogenetic tree based on ORF470 sequences. Numbers in the branches represent Bayesian posterior probability and bootstrap value. The three collapsed groups cluster *Eimeria* sequences from *M. musculus* of this study. Reference sequences from other rodents were included. The scale bar represents sequence divergence. Hosts for closely related sequences of *E. falciformis* and *E. vermiformis* are indicated in the expanded form of the group. * Represent sequences generated in the present study. Tissue of origin is indicated in brackets. Sequences in bold were included in the multi-marker phylogenetic inference.

### 3.4.2 Low genetic diversity of mt COI in rodent *Eimeria* isolates

With the aim to estimate the genetic diversity of isolates of *Eimeria* from different rodent hosts, we constructed a haplotype network (Figure 3.5) from 161 COI sequences obtained in this study combined with 59 previously published sequences (alignment of 459 bp without gaps). The network comprised 20 different haplotypes with up to 14 polymorphic nucleotide sites among them. The network confirms the lack of genetic differentiation of *E. falciformis* and *E. vermiformis* from some isolates described as *E. apionodes* in non-*Mus* hosts using sequences of COI.

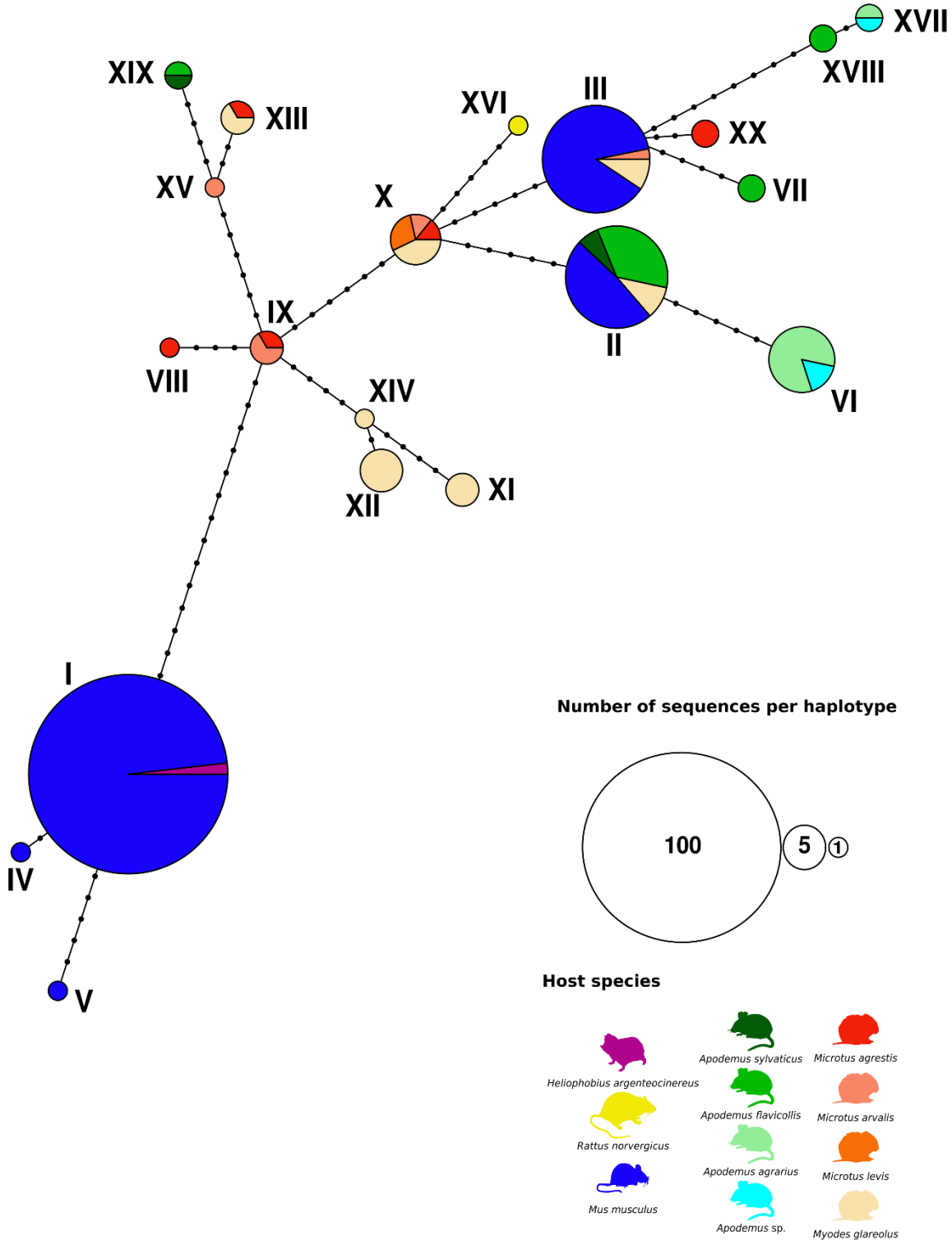



**Figure 3.5 Statistical parsimony network of *Eimeria* spp. haplotypes for COI sequences.** Network based on a 459 bp region of the gene coding for the mitochondrial cytochrome *c* oxidase from *Eimeria* isolates detected in rodents (*Mus musculus, Apodemus flavicollis, A. sylvaticus, A. agrarius*) caught in Europe. Previously published sequences from different species of Eimeria infecting cricetid and murid rodents were also included. Colouring of each haplotype is based on the host species from the *Eimeria* isolate. Every haplotype is marked with a consecutive number and its size indicates the number of sequences included on it. Each node represents a mutational step between two haplotypes.

### 3.4.3 Multi-locus genotyping

To determine whether markers with a higher resolution could distinguish host-usage patterns for the “rodent parasite models” *E. falciformis* and *E. vermiformis*, we designed a multi-locus sequence typing approach. 35 markers targeting exons in the nuclear genome (Supplementary data S3.4) and 5 regions of the apicoplast genome were amplified for 19 samples from *Apodemus* spp. hosts, 12 samples from house mice and corresponding regions from the reference genome of *E. falciformis* and *E. vermiformis* were included. All the isolates used correspond to *Eimeria* species with different morphology (Supplementary data S3.10).

A multivariate analysis identified three clusters of isolates for the nuclear markers: one group included the laboratory isolate of *E. vermiformis,* another the isolate of *E. falciformis* and a third group only contained *Eimeria* isolates from *Apodemus agrarius* (Figure 3.6B). This result was corroborated by phylogenetic analysis of SNPs (2019 informative alignment columns). We excluded prevalent indels from this analysis. Indels in protein coding genes (all “in-frame” with a length divisible by three) correspond to homopolymeric amino acid repeats (HAARs) and are expected in protein coding genes of *Eimeria* spp.[143,245] (supplementary data S8). Three clades were recovered in this tree (Figure 3.6A): Despite the apparently morphological differences (Supplementary data S3.10, S3.11), the laboratory isolate of *E. vermiformis* from *Mus musculus* was indistinguishable from a field isolate from *Apodemus flavicollis*. Other isolates from house mouse and *A. flavicollis* and *A. sylvaticus* clustered in an unresolved internal relationship with house mouse *E. vermiformis* isolates (Group II). We note that four of five sequences for *E. vermiformis* from house mice were amplified from ileum tissue, the primary location of infection with this species (in contrast, *E. falciformis* infects primarily the caecum[233]). A second clade recovered by nuclear multi-locus analysis contained *E. falciformis* from house mice. This clade showed a well-supported substructure in which 7 house mouse field isolates grouped with the laboratory isolate BayerHaberkorn1970 but were separated from 4 *Eimeria* isolates from *A. flavicollis* (Group III). This substructure agrees with the morphological difference previously observed based on the presence of polar granule in *E. falciformis* and its absence in *E. apionodes* (Supplementary data S3.10, S3.11).

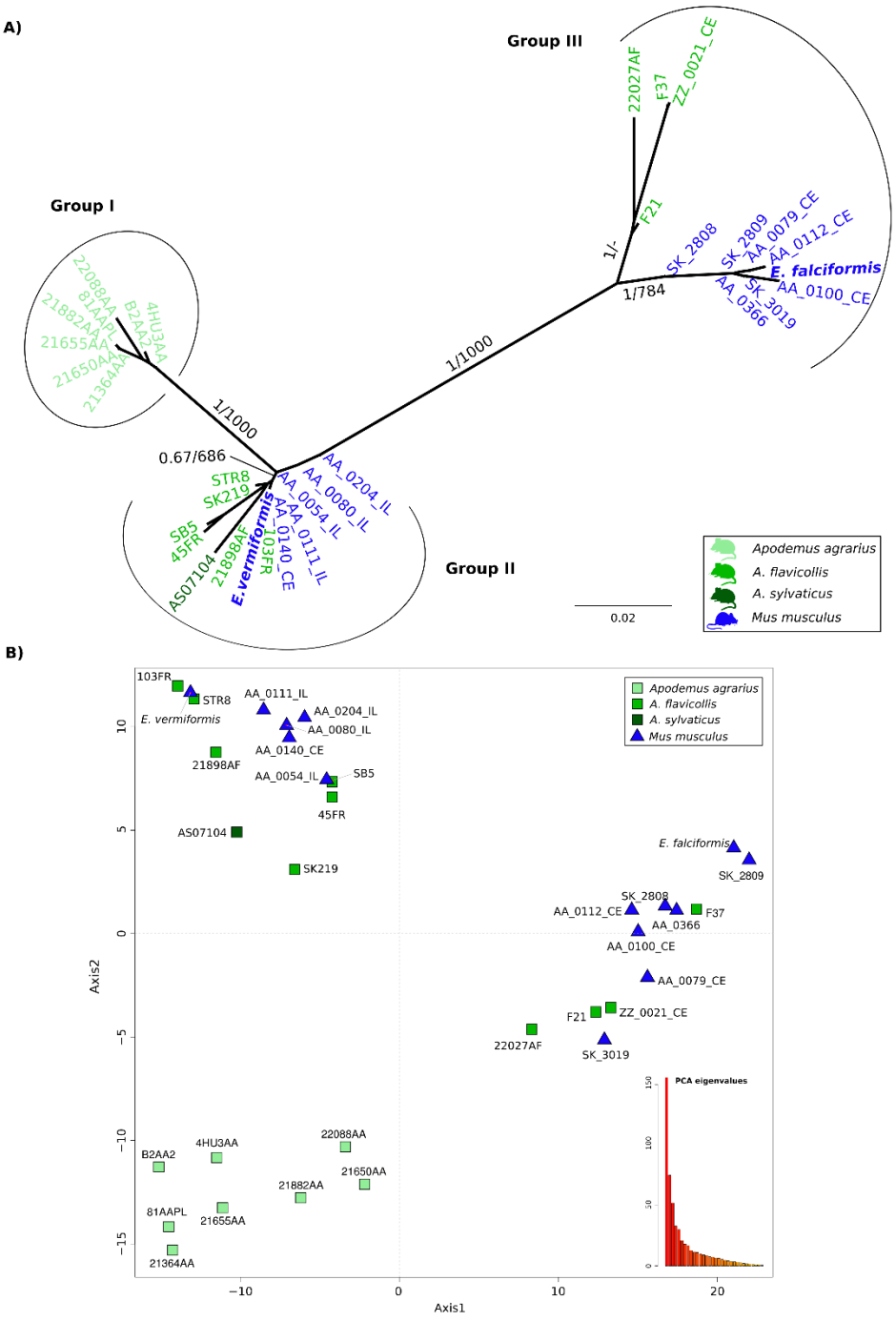



**Figure 3.6 Nuclear multilocus genotyping of *Eimeria* isolates from *Mus musculus* and *Apodemus*.** A) The phylogenetic tree was estimated with a multi-marker dataset formed with 35 nuclear markers from 31 *Eimeria* isolates derived from wild *Mus musculus* and three species of *Apodemus* (*A. agrarius*, *A. sylvaticus*, *A. flavicollis*). *Eimeria falciformis* and *E. vermiformis* sequences were included as reference. The scale bar represents sequence divergence. Colour represents the host of origin for the isolates. Bootstrap support values and Bayesian posterior probabilities are shown on branches. B) Principal component analysis based on single nucleotide polymorphisms (SNPs) from the same *Eimeria* isolates. Samples form three clusters. Shape indicates the genus of host and colours the species. Eigenvalues of the dimensions are shown in an insert to visualize the proportion of variance explained by the axes.

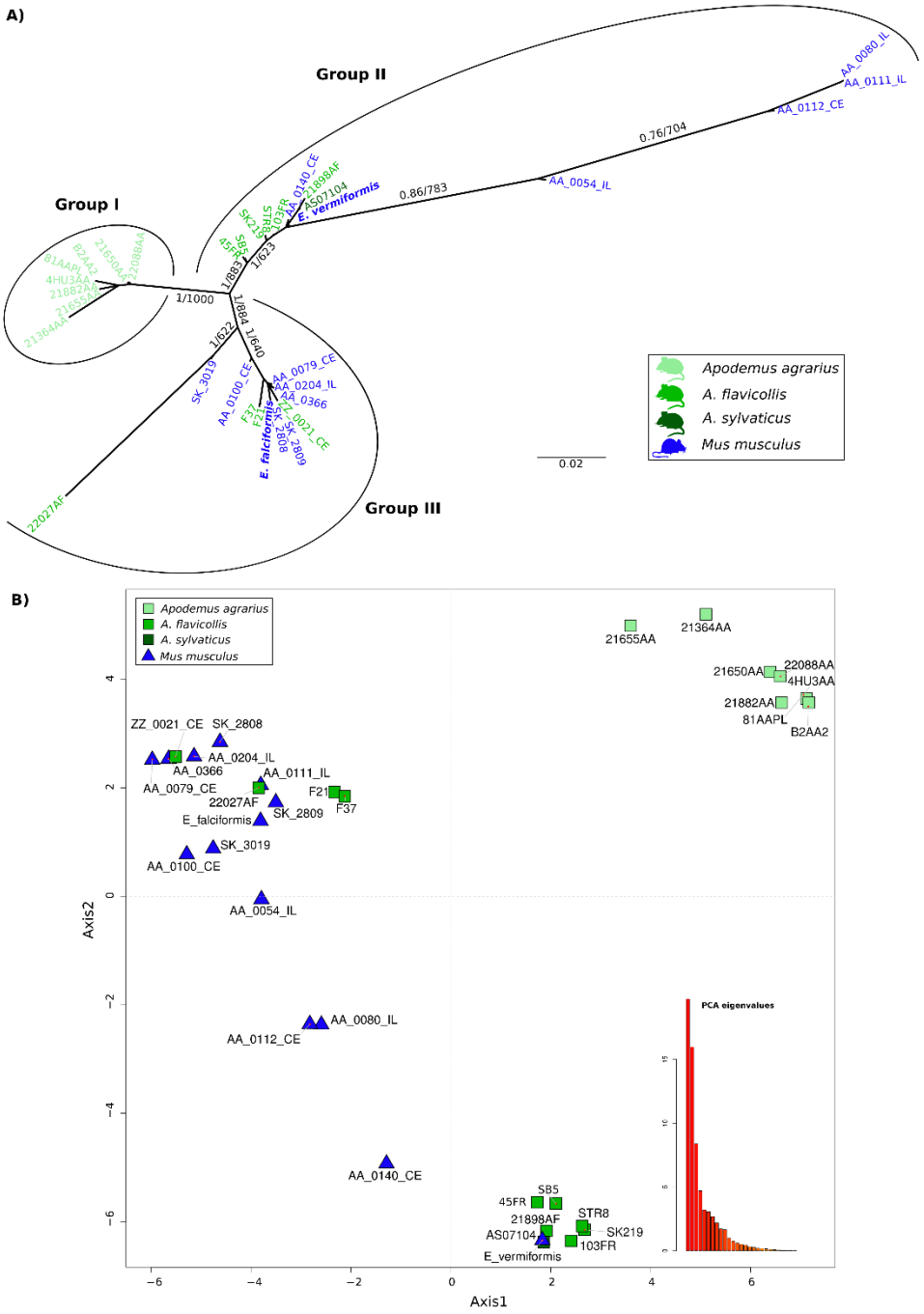



**Figure 3.7 Apicoplast multilocus genotyping of *Eimeria* isolates from *Mus musculus* and *Apodemus*.** A) The phylogenetic tree was estimated with a multi-marker dataset formed with 5 apicoplast markers from 31 *Eimeria* wild isolates derived from *Mus musculus* and three species of *Apodemus* (*A. agrarius*, *A. sylvaticus*, *A. flavicollis*). *Eimeria falciformis* and *E. vermiformis* sequences were included as reference. The scale bar represents the sequence divergence. Colour represents the host of origin for the isolates. Bootstrap support values and Bayesian posterior probabilities are shown on branches. B) Principal component analysis based on single nucleotide polymorphisms (SNPs) from the same *Eimeria* isolates. Samples form three clusters based on the similarities for all the SNPs. Shape indicates the genus of host and colours the species. Eigenvalues of the dimensions are shown in an insert to visualize the proportion of variance explained by the axes.

Analyses based on apicoplast markers (both multivariate clustering and phylogenetic analyses; Figure 3.7) identified similar groups: a well separated cluster with isolates from *A. agrarius,* a cluster containing *E. vermiformis* and another containing *E. falciformis* isolates. Some differences between the apicoplast and nuclear markers were obvious, though. *Eimeria* isolates from *M. musculus* (AA_0054_IL, AA_0080_IL, AA_0111_IL and AA_0112_CE) were less similar to the *E. vermiformis* group, leading to a multivariate clustering between the *E. falciformis* and *E. vermiformis* groups (Figure 3.6B). This was recovered in a phylogenetic tree as isolates appeared at the end of a long branch in the *E. vermiformis* group (Figure 3.7A). In an analysis of apicoplast markers, the *E. falciformis* isolates from *Mus* were not differentiated from those from *A. flavicollis*. Inspection of phylogenetic trees for individual markers (Supplementary data S3.12) highlighted problems with the apicoplast dataset: samples that had been previously reported as co-infected with *E. ferrisi* (AA_0080_IL, AA_0111_IL and AA_0112_CE), showed an aberrant clustering for different markers. Samples AA_0080_IL and AA_0111_IL clustered in the group of *E. falciformis* with Ap12, while AA_0112_CE clustered with Ap5, in disagreement with the consensus species trees for other markers. We conclude that for these samples *E. ferrisi* or even *E. falciformis* apicoplast sequences were likely amplified and recovered as the majority sequence.

## 3.5 Discussion

We studied whether host specificity of Coccidia can be assessed with currently used molecular markers, using the example of *Eimeria* species in house mice and related rodents. We found that commonly used phylogenetic markers, nu 18S rRNA and mt COI, are not sufficiently variable to differentiate parasite isolates that would be regarded as separate species based on host usage. The relatively rarely used marker ap ORF470 from the apicoplast genome seems to provide slightly better resolution. We developed a multi-locus genotyping approach to show that *E. falciformis* from the house mouse can likely be distinguished from related isolates from other hosts based on nuclear markers. In contrast, even with this high-resolution approach *E. vermiformis* from house mice and isolates from other host species were found in a nested and unresolved cluster.

Phylogenies derived from each of the analysed markers (esp. 18S) confirmed the topology of rodent *Eimeria* species observed before at deeper nodes of the phylogeny[78,131,132]. At the tips of the phylogeny, 18S sequences of *E. falciformis* and *E. vermiformis* isolates clustered with isolates from hosts of different genera or even families (Figure 3.2). This result was expected to some extent, as phylogenetic analyses with 18S sequences usually fail to separate closely related parasites isolated from closely related hosts[78].

Previous studies described COI as a universal barcode variable enough to resolve relationships between coccidians, including *Eimeria* [72,78]. We therefore expected to differentiate our house mouse isolates from species found in other hosts using COI. Neither phylogenetic (Figure 3.3) nor haplotype inference (Figure 3.5), however, supported differentiation of *E. falciformis* and *E. vermiformis* from some of the isolates described as *E. apionodes.* Many of the COI sequences were even identical for isolates from different hosts. Limited resolution of COI outside of metazoans has been reported before[246]. Rodent hosts of *Eimeria,* in the families Muridae (*Mus, Rattus, Apodemus*) and Cricetidae (*Myodes, Microtus*)*,* diverged around 25 Million years ago[247,248] and it seems possible that COI of Coccidia evolves at such slow rates that it fails to differentiate *Eimeria* species with similar divergence. We stress that for rodent Coccidia, COI should not be assumed to resolve *bona fide* species with different host usage.

The potential of the apicoplast marker ORF470 to distinguish rodent *Eimeria* species has been highlighted before[77,131], but few studies have followed the recommendation to use this marker. Consequently, few database sequences are available. Phylogenetic analysis of these sequences (Figure 3.4) separates our three species clusters well and shows hints of internal structure separating *E. apionodes* derived from *A. flavicollis* from house mouse isolates. Our work increases the number of sequences available for ORF470 and supports its use as a marker for discrimination of *Eimeria* species.

To test host specificity for *E. falciformis*, *E. vermiformis* (from house mice) and *E. apionodes* (from *Apodemus* spp.), we established and used a multi-locus sequence typing protocol. Our multi-locus approach supports a differentiation of *E. falciformis* (infecting the house mouse[138,168]) from *E. apionodes* (infecting *A. flavicolis*[232]). The same approach was unable to distinguish *M. musculus* derived *E. vermiformis* isolates from one "*E. apionodes"* isolate from *A. flavicollis* (Figures 3.2, 3.3, 3.4, 3.6 and 3.7). This suggests a broad host range of genetically indistinguishable *Eimeria* isolates which have been assigned to paraphlyetic species *E. apionodes* and *E. vermiformis*.

Multi-locus genotyping using apicoplast markers showed some discrepancies with the nuclear analysis. These discrepancies can be attributed to double infections previously discovered in those particular isolates[233]. Compared to the nuclear genome, the apicoplast genome is present in much higher copy numbers[143]. This, combined with more conserved primer binding sites, can lead to amplification of non-target sequences such as those of the prevalent *E. ferrisi*[233] creating artificial "chimeric" isolates in case of double infections.

We use our system also as a test case whether the commonly used markers (18S, COI) provide enough resolution to assess parasite specificity. We conclude that unresolved genetic clusters

and monomorphic haplotypes currently identified via 18S and COI genotyping should not be assumed to indicate parasite species with generalist host usage. Novel nuclear markers are needed in addition to ORF470 to analyse host species specificity of rodent *Eimeria.* Care must be taken to avoid potential artifacts introduced by double infection and mixed amplification.

Whether other *Eimeria* species from different rodent hosts are indeed phylogenetically distinguishable species (or whether genetically differentiable clusters show different host usage) is still an open question. This question needs to be addressed more broadly with markers providing higher resolution than 18S or COI. This question is highly relevant as hypotheses, assumptions and predictions concerning host-parasite interactions from evolutionary[95,98,249,250], ecological[251–253] and mechanistic[254] perspectives depend on the placement of parasite species in the specialist-generalist continuum[95].

# Chapter 4. General discussion

## 4.1 Summary of the thesis

This thesis aimed to examine different approaches for parasite species identification. The work focused on describing parasite diversity in natural systems and testing whether species identification can rely on assumptions of host specificity. Specifically, morphology, single-marker and multi-locus genotyping and phylogenetic methods were implemented to disentangle the challenging taxonomy of the genus *Eimeria* and its host association with different rodent species.

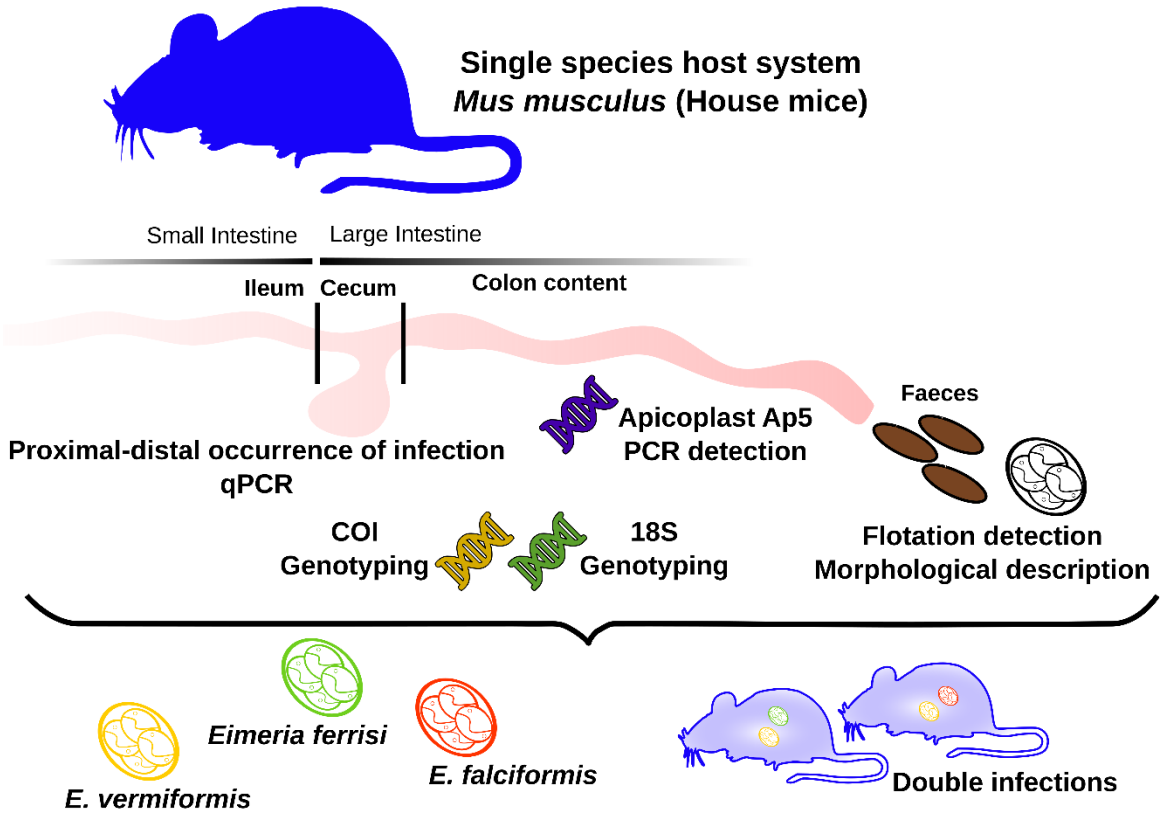



**Figure 4.1 Multi-method strategy to assess *Eimeria* detection and identification at species level.** The approach involved both morphological and molecular methods, allowing three different *Eimeria* sp. to be identified, as well as detection of double infections.

The high dependency of *Eimeria* identification on morphological features and host association was broadly addressed in Chapter 1. In **Chapter 2**, *Eimeria* parasites were identified in wild populations of house mice (*Mus musculus*). The use of a single species host system allowed me to focus on the challenges of morphological identification. Hence it was feasible to provide

complementary tools to achieve reliable species identification. The molecular tools implemented here included a highly sensitive detection PCR utilising a novel apicoplast-specific primer pair, genotyping PCRs with established markers for species identification (18S and COI) and a novel qPCR strategy for niche approximation and relative quantification of the parasite. With this joint strategy, it was possible to not only identify three different *Eimeria* species in the house mouse (*E. ferrisi, E. falciformis and E. vermiformis*), but also to detect double infections, which has not been reported until now for this particular host (Figure 4.1). The proposed multi-method strategy represents a promising approach to assess coccidia at the species level. Thus, test hypotheses in ecology, evolution and epidemiology of these parasites can benefit from incorporating morphological, molecular and life-cycle characteristics simultaneously in studies with wildlife populations.

**Chapter 3** expanded the results from the previous Chapter and aimed to test the host association based on the phylogeny of rodent-infecting *Eimeria*. Here, SSU 18S rRNA and COI showed insufficient resolution to confidently distinguish between different *Eimeria* species from Murid and Cricetid rodents. Moreover, phylogeny inferred with these *a priori* well-established markers did not agree with the host association of these parasites. Therefore, 18S and COI phylogeny should not be used to assume low host specificity of parasite species. Through the multi-locus genotyping established here, it was possible to differentiate phylogenetic clusters unresolved by 18S and COI. On one hand, the implementation of this multi-locus genotyping allowed the differentiation of *E. falciformis/E. apionodes* isolates and linked identification to their host usage. On the other hand, it revealed the inability to distinguish isolates clustered together from *E. vermiformis* infecting *M. musculus* and *E. apionodes* infecting *Apodemus flavicollis*. The latter result indicates a broad host range of these indistinguishable isolates (Figure 4.2). Thus, the phylogenetic specificity suggested for *Apodemus*-infecting *Eimeria*[91] is disputed.

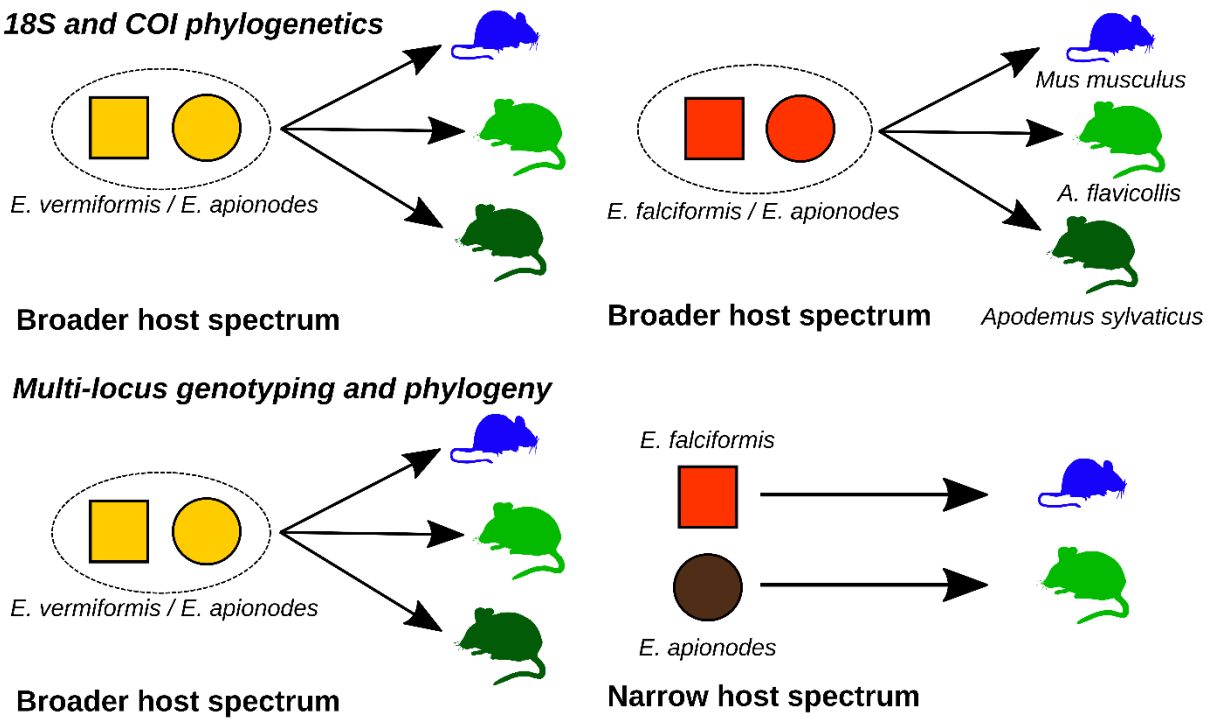



**Figure 4.2 Comparison between single-marker and multi-locus genotyping.** Multi-locus genotyping resolved *E. falciformis*/*E. apionodes* clustering while 18S and COI did not resolve this relationship. The resolved species showed a narrow host spectrum and suggest potential infection only of one host species. Neither 18S/COI nor multi-locus genotyping resolved *E. vermiformis*/*E. apionodes* clustering, suggesting a broader host spectrum.

## 4.2 The Species conflict in *Eimeria*

As explained in the introduction no species identification method provides an unambiguous way to delineate all-natural groups, as taxa may vary in their degree of morphological dissimilarity, amount of gene flow with related taxa, or adaptation to different niches[6]. However, the species identification methods are essential to make species concepts useful.

The MSC requires methods to define morphological distinctiveness among organisms and, simultaneously, to confirm consistency and persistence of those morphological differences. In general, the consistency and persistence is hardly addressed for any organism or isolate but it is particularly challenging for parasites. Here, I argue that especially an inclusion of host usage in an identification approach disregards the “consistency and persistence” of morphological characters demanded by the concept. The host environment is declared a distinguishing feature instead of testing its effect on morphological characterization. In contrast, the PSC allows a “simple” identification, sustained by methods that almost overlap entirely with the definition of the concept. The overall results of this thesis confirmed that the failure in

*Eimeria* species identification is linked to the limitations in current conventional methodologies. The phylogenetic methods not only provided a supported identification of *Eimeria* in *M. musculus,* but also distinguished closely related species and allowed to test hypotheses on their host usage.

Previous studies indicated that polymorphism in *Eimeria* oocyst from birds and rodents is common and complicates the identification based on diagnostic characters[126,127]. For that reason, during **Chapter 2** of this study, *M. musculus* was used as a single host system to abate the problem of host-mediated phenotypic and morphological variability between isolates. Considering that the original identification of *E. falciformis, E. ferrisi* and *E. vermiformis* in *M. musculus* was based solely on morphological traits, the authors indicated clear morphological distinctiveness among them[112]. According to the results shown in **Chapter 2**, the three *Eimeria* species did not exhibit considerable phenotypic variability evidenced by a noticeable overlap of the morphological diagnostic characters. In contrast, the phylogenetic clustering showed a consistent and highly-supported monophyletic pattern, complying with the PSC. Not only does this result highlight a robust phylogenetic position between *Mus*-infecting *Eimeria* species, but also provides consistency to the identification based on morphological traits. Collectively, incorporating phylogeny is crucial to improve confidence in morphological species assignments in coccidia from wildlife samples and should not be disregarded.

Phylogenetic methods to delineate species require a precise selection of genetic markers that represent both the historical relatedness well and possess enough character variability to distinguish the organism. The general assumption regarding the use of SSU 18S rRNA as a feasible marker for *Eimeria* and coccidia identification and phylogenetic inference[206,231,255,256] has been questioned before based on the findings of paralogues within the genome of poultry-infecting species[257,258]. Regardless, this marker continues to be the most represented genetic marker for this group of parasites in public databases. According to the results presented in **Chapter 2**, SSU 18S rRNA presented enough resolution to distinguish *Mus musculus*-infecting *Eimeria* species. However, **Chapter 3** results indicate that 18S was unable to resolve closely related isolates from different host origins. In the same chapter, COI usage as “universal” genetic barcode for species was demystified. Previous studies on *Eimeria* suggested COI as a suitable marker based on its sufficient DNA variability, multi-copy presence, sufficient divergence from host DNA, low intraspecific variation and being free of introns due to its mitochondrial origin[72]. Here, rodent-infecting *Eimeria* has shown identical COI sequences. In contrast, the apicoplast-derived and underrepresented marker ORF470[131] provided congruent results with the multi-locus genotyping, similarly exhibiting higher resolution compared to the

standard markers. Therefore, the usage of ORF470 as an alternative marker is encouraged for further studies on coccidia and to diminish its underrepresentation by incorporating a vast number of new sequences to the public databases.

In 1984, Long and Joyner[124] recognised the problems in *Eimeria* species identification. The authors referred to the BSC to emphasise that current identification approaches related to the MSC should not be solely used to sustain Coccidia species in general, and particularly for *Eimeria* species. The authors additionally proposed a series of parameters required to overcome the problem in *Eimeria* species delineation. However, it is not clear how it could be possible to demonstrate genetic recombination and reproductive independence between two strains of the same species of *Eimeria*, especially for those isolates collected from wild hosts. In that sense, the phylogenetic methods for species identification presented in this thesis represent an alternative to morphological assessments and those approaches that attempt to adopt the BSC as reference. While 18S or COI as single or concatenated datasets did not provide sufficient resolution to distinguish *E. falciformis* from *E. apionodes.* The multi-locus genotyping for *Eimeria* provided an extensive set of phylogenetic markers to address the evolutionary history and identification of *Eimeria* spp. with higher support than the conventional single-marker approaches [259].

Considering the genetic similarity revealed by the multi-locus genotyping between *E. vermiformis* and *E. apionodes* in **Chapter 3**, the re-identification of these *Eimeria* species and subsequent unification into single species may be required. Re-descriptions and re-assignments previously made for Coccidia mainly justified morphological similarities between isolates from different hosts[222,260–262] and few have been founded on genotyping data[263]. Therefore, a full genomic comparison of these isolates might represent additional evidence to define whether their taxonomy should be revised. Independent of taxonomy, biological questions can be asked based on the approaches developed here.

## 4.3 Conclusion

This thesis emphasised the necessity to uniform and combine strategies in *Eimeria* detection,

quantification, and identification. Beyond the particular focus of this work in rodent *Eimeria*, uniform identification approaches are essential to promote a more reliable understanding of *Eimeria* spp. from an evolutionary and ecological level. In conclusion, phylogenetic identification of *Eimeria* from rodents represents a preferable way to distinguish closely related isolates with indistinguishable morphology. This thesis provides the molecular approaches to identify *Eimeria* achieving species level and allow to ask questions on host usage and specificity based on isolates from natural systems.

# References


1. Mayden, R. L. A hierarchy of species concepts: the denouement in the saga of the species problem. in *Species. The units of biodiversity.* (eds. Claridge, M. F., Dawah, H. A. & Wilson, M. R.) 381–423 (Chapman & Hall, 1997).
2. Cracraft, J. Species concept in theoretical and applied biology: a systematic debate with consequences. in *Species concepts and phylogenetic theory: a debate* (eds. Wheeler, Q. D. & Meier, R.) 3–14 (Columbia University Press, 2000).
3. Mayr, E. Birds collected during the Whitney South Sea Expedition. XLVIII. Notes on the Polynesian species of Aplonis. *Am. Museum Novit.* **9**, 1–6 (1942).
4. Mayr, E. *Animal species and evolution*. (Harvard University Press, 1963).
5. Mayr, E. *Populations, Species, and Evolution: An abridgment of animal species and evolution.* (Harvard University Press, 1970).
6. Bromham, L. *An introduction to molecular evolution and phylogenetics*. (Oxford University Press, 2016).
7. Hey, J. On the failure of modern species concepts. *Trends Ecol. Evol.* **21**, 447–450 (2006).
8. Cronquist, A. Once again, what is a species? in *BioSystematics in Agriculture* (ed. Romberger, J. A.) 3–20 (Allenheld, Osman & Company, 1978).
9. Strauss, R. E. & Bond, C. E. Taxonomic Methods : Morphology. in *Methods for Fish Biology* (eds. Schreck, C. B. & Moyle, P. B.) 109–140 (American Fisheries Society, 1990).
10. Foote, M. Sampling, Taxonomic Description, and our Evolving Knowledge of Morphological Diversity. *Paleobiology* **23**, 181–206 (1997).
11. Armenteros, M., Ruiz-Abierno, A., Vincx, M. & Decraemer, W. A morphometric analysis of the genus Terschellingia (Nematoda: Linhomoeidae) with redefinition of the genus and key to the species. *J. Mar. Biol. Assoc. United Kingdom* **89**, 1257–1267 (2009).
12. Thompson, M. J. & Timmermans, M. J. T. N. Characterising the phenotypic diversity of *Papilio dardanus* wing patterns using an extensive museum collection. *PLoS One* **9**, 1–9 (2014).
13. Töpfer, T. Morphological Variation in Birds: Plasticity, Adaptation, and Speciation. in *Bird Species: How they arise, modify and vanish* (ed. Tietze, D. T.) 63–74 (Springer Open, 2018).
14. Minelli, A. The status of taxonomic literature. *Trends Ecol. Evol.* **18**, 75–76 (2003).
15. Pearson, D. L., Hamilton, A. L. & Erwin, T. L. Recovery plan for the endangered taxonomy profession. *Bioscience* **61**, 58–63 (2011).
16. Cracraft, J. The Significance of Phylogenetic Classifications for Systematic and Evolutionary Biology. in *Numerical Taxonomy* (ed. Felsenstein, J.) 1–17 (Springer-

Verlag, 1983).

17. Mishler, B. D. & Brandon, R. N. Individuality, pluralism, and the phylogenetic species concept. *Biol. Philos.* **2**, 397–414 (1987).

18. McKitrick, M. & Zink, R. Species concept in ornithology. *Condor* **90**, 1–14 (1988).

19. Nixon, K. C. & Wheeler, Q. D. An Amplification of the Phylogenetic Species Concept. *Cladistics* **6**, 211–223 (1990).

20. Sites, J. W. & Marshall, J. C. Delimiting species: A Renaissance issue in systematic biology. *Trends Ecol. Evol.* **18**, 462–470 (2003).

21. Mark Pagel. Inferring historical patterns of biological evolution. *Nature* **401**, 877–884 (1999).

22. Edman, J. C. *et al.* Ribosomal RNA sequence shows *Pneumocystis carinii* to be a member of the Fungi. *Nature* **334**, 519–522 (1988).

23. Wolters, J. The troublesome parasites - molecular and morphological evidence that Apicomplexa belong to the dinoflagellate-ciliate clade. *BioSystems* **25**, 75–83 (1991).

24. Fischer, W. M. & Palmer, J. D. Evidence from small-subunit ribosomal RNA sequences for a fungal origin of *Microsporidia*. *Mol. Phylogenet. Evol.* **36**, 606–622 (2005).

25. Brabec, J., Waeschenbach, A., Scholz, T., Littlewood, D. T. J. & Kuchta, R. Molecular phylogeny of the Bothriocephalidea (Cestoda): Molecular data challenge morphological classification. *Int. J. Parasitol.* **45**, 761–771 (2015).

26. Fišer, C., Robinson, C. T. & Malard, F. Cryptic species as a window into the paradigm shift of the species concept. *Mol. Ecol.* **27**, 613–635 (2018).

27. Gilmour, J. S. L. A taxonomic problem. *Nature* **139**, 1040–1042 (1937).

28. Lidén, M. §193 and §246 of Philosophia botanica do not support that Linnaeus was a “typologist”. *Taxon* **69**, 213–216 (2020).

29. Dayrat, B. Towards integrative taxonomy. *Biol. J. Linn. Soc.* **85**, 407–415 (2005).

30. Goldstein, P. Z. & DeSalle, R. Integrating DNA barcode data and taxonomic practice: Determination, discovery, and description. *BioEssays* **33**, 135–147 (2011).

31. Pante, E., Schoelinck, C. & Puillandre, N. From integrative taxonomy to species description: One step beyond. *Syst. Biol.* **64**, 152–160 (2015).

32. Hebert, P. D. N. N., Cywinska, A., Ball, S. L. & DeWaard, J. R. Biological identifications through DNA barcodes. *Proc. R. Soc. B Biol. Sci.* **270**, 313–321 (2003).

33. Schoch, C. L. *et al.* Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 6241–6246 (2012).

34. Hajibabaei, M., Singer, G. A. C. & Hickey, D. A. Benchmarking DNA barcodes: An assessment using available primate sequences. *Genome* **49**, 851–854 (2006).

35. Hajibabaei, M., Singer, G. A. C., Hebert, P. D. N. & Hickey, D. A. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends Genet.* **23**, 167–172 (2007).

36. Hoberg, E. P. Foundations for an integrative parasitology: Collections, archives, and biodiversity informatics. *Comp. Parasitol.* **69**, 124–131 (2002).

37. Blaxter, M. Parasite genomics. in *Molecular Medical Parasitology* (eds. Marr, J., Nielsen, T. W. & Komuniecki, R. W.) 3–27 (Elsevier, 2003).

38. Powers, T. Nematode Molecular Diagnostics: From bands to barcodes. *Annu. Rev. Phytopathol.* **42**, 367–383 (2004).

39. Walker, G., Dorrell, R. G., Schlacht, A. & Dacks, J. B. Eukaryotic systematics: A user's guide for cell biologists and parasitologists. *Parasitology* **138**, 1638–1663 (2011).

40. De Meeûs, T. & Renaud, F. Parasites within the new phylogeny of eukaryotes. *Trends Parasitol.* **18**, 247–250 (2002).

41. Poulin, R. & Morand, S. The diversity of parasites. *Q. Rev. Biol.* **75**, 277–293 (2000).

42. Poulin, R. & Morand, S. *Parasite Biodiversity*. (Smithsonian Institution, 2004).

43. Brazenor, A. K., Saunders, R. J., Miller, T. L. & Hutson, K. S. Morphological variation in the cosmopolitan fish parasite *Neobenedenia girellae* (Capsalidae: Monogenea). *Int. J. Parasitol.* **48**, 125–134 (2018).

44. Monis, P. T., Andrews, R. H., Mayrhofer, G. & Ey, P. L. Genetic diversity within the morphological species *Giardia intestinalis* and its relationship to host origin. *Infect. Genet. Evol.* **3**, 29–38 (2003).

45. Monis, P. T., Caccio, S. M. & Thompson, R. C. A. Variation in *Giardia*: towards a taxonomic revision of the genus. *Trends Parasitol.* **25**, 93–100 (2009).

46. Honigberg, B. A Morphological Abnormality in the Cestode, *Dipylidium caninum*. *Traactions Am. Microsc. Soc.* **63**, 342–344 (1944).

47. Beugnet, F., Labuschagne, M., Vos, C. De, Crafford, D. & Fourie, J. Analysis of *Dipylidium caninum* tapeworms from dogs and cats, or their respective fleas: Part 2. Distinct canine and feline host association with two different Dipylidium caninum genotypes. *Parasite* **25**, (2018).

48. Woese, C. R. Interpreting the universal phylogenetic tree. *Proc. Natl. Acad. Sci.* **97**, 8392–8396 (2000).

49. Dacks, J. B., Walker, G. & Field, M. C. Implications of the new eukaryotic systematics for parasitologists. *Parasitol. Int.* **57**, 97–104 (2008).

50. Nadler, S. A. Molecular approaches to studying Helminth population genetics and phylogeny. *Int. J. Parasitol.* **20**, 11–29 (1990).

51. Sogin, M. L. & Silberman, J. D. Evolution of the protists and protistan parasites from the perspective of molecular systematics. *Int. J. Parasitol.* **28**, 11–20 (1998).

52. Cavalier-Smith, T. *Kingdom protozoa and its 18 phyla*. *Microbiological Reviews* **57**, (1993).

53. Blaxter, M. L. *et al.* A molecular evolutionary framework for the phylum Nematoda. *Nature* **392**, 71–75 (1998).

54. Nadler, S. A. *et al.* Molecular phylogeny of clade III nematodes reveals multiple origins of tissue parasitism. *Parasitology* **134**, 1421–1442 (2007).

55. Morrison, D. A. Evolution of the Apicomplexa: where are we now? *Trends Parasitol.* **25**, 375–382 (2009).

56. Goggin, C. L. & Barker, S. C. Phylogenetic position of the genus *Perkinsus* (Protista, Apicomplexa) based on small subunit ribosomal RNA. *Mol. Biochem. Parasitol.* **60**, 65–70 (1993).

57. Baverstock, P. R., Fielke, R., Johnson, A. M., Bray, R. A. & Beveridge, I. Conflicting phylogenetic hypotheses for the parasitic platyhelminths tested by partial sequencing of 18S ribosomal RNA. *Int. J. Parasitol.* **21**, 329–339 (1991).

58. Lin, D., Hanson, L. A. & Pote, L. M. Small Subunit Ribosomal RNA Sequence of *Henneguya exilis* (Class Myxosporea) identifies the Actinosporean stage from an Oligochaete host. *J. Euk. Microbiol* **46**, 66–68 (1999).

59. Holmdahl, O. J. M., Morrison, D. A., Ellis, J. T. & Huong, L. T. T. Evolution of Ruminant *Sarcocystis* (Sporozoa) Parasites Based on Small Subunit rDNA Sequences. *Mol. Phylogenet. Evol.* **11**, 27–37 (1999).

60. Kent, M. L., Khattra, J., Hedrick, R. P. & Devlin, R. H. *Tetracapsula renicola* n. sp. (Myxozoa: Saccosporidae); the PKX myxozoan - The cause of proliferative kidney disease of salmonid fishes. *J. Parasitol.* **86**, 103–111 (2000).

61. Hillis, D. M., Moritz, C., Porter, C. A. & Baker, R. J. Evidence for biased gene conversion in concerted evolution of ribosomal DNA. *Science (80-. ).* **251**, 308–310 (1991).

62. Hillis, David.M., Dixon, M. T. Ribosomal DNA: Molecular Evolution and Phylogenetic Inference. *Q. Rev. Biol.* **66**, (1991).

63. Ganley, A. R. D. & Kobayashi, T. Highly efficient concerted evolution in the ribosomal DNA repeats: Total rDNA repeat variation revealed by whole-genome shotgun sequence data. *Genome Res.* **17**, 184–191 (2007).

64. Dorris, M., De Ley, P. & Blaxter, M. Molecular analysis of nematode diversity and the evolution of parasitism. *Parasitol. Today* **15**, 188–193 (1999).

65. De Ley, P. *et al.* An integrated approach to fast and informative morphological vouchering of nematodes for applications in molecular barcoding. *Philos. Trans. R. Soc. B Biol. Sci.* **360**, 1945–1958 (2005).

66. Rooney, A. P. Mechanisms underlying the evolution and maintenance of functionally heterogeneous 18S rRNA genes in apicomplexans. *Mol. Biol. Evol.* **21**, 1704–1711 (2004).

67. Barta, J. R., Ogedengbe, J. D., Martin, D. S. & Smith, T. G. Phylogenetic position of the adeleorinid coccidia (myzozoa, Apicomplexa, coccidia, eucoccidiorida, Adeleorina) inferred using 18s rDNA sequences. *J. Eukaryot. Microbiol.* **59**, 171–180 (2012).

68. Blaxter, M. L. The promise of a DNA taxonomy. *Philos. Trans. R. Soc. B Biol. Sci.* **359**, 669–679 (2004).

69. Homan, W. L. *et al.* Comparison of the internal transcribed spacer, ITS 1, from *Toxoplasma gondii* isolates and *Neospora caninum*. *Parasitol. Res.* **83**, 285–289 (1997).

70. Šlapeta, J. R. *et al.* Dog shedding oocysts of *Neospora caninum*: PCR diagnosis and molecular phylogenetic approach. *Vet. Parasitol.* **109**, 157–167 (2002).

71. Casiraghi, M., Anderson, T. J. C., Bandi, C., Bazzocchi, C. & Genchi, C. A phylogenetic analysis of filarial nematodes: Comparison with the phylogeny of *Wolbachia* endosymbionts. *Parasitology* **122**, 93–103 (2001).

72. Ogedengbe, J. D., Hanner, R. H. & Barta, J. R. DNA barcoding identifies *Eimeria* species and contributes to the phylogenetics of coccidian parasites (Eimeriorina, Apicomplexa, Alveolata). *Int. J. Parasitol.* **41**, 843–850 (2011).

73. Galen, S. C. *et al.* The polyphyly of *Plasmodium*: Comprehensive phylogenetic analyses of the malaria parasites (Order Haemosporida) reveal widespread taxonomic conflict. *R. Soc. Open Sci.* **5**, 171780 (2018).

74. Bachvaroff, T. R., Handy, S. M., Place, A. R. & Delwiche, C. F. Alveolate Phylogeny Inferred using Concatenated Ribosomal Proteins. *J. Eukaryot. Microbiol.* **58**, 223–233 (2011).

75. Callejón, R. *et al.* Molecular characterization and phylogeny of whipworm nematodes inferred from DNA sequences of cox1 mtDNA and 18S rDNA. *Parasitol. Res.* **112**, 3933–3949 (2013).

76. Schaer, J. *et al.* Phylogeny of *Hepatocystis* parasites of Australian flying foxes reveals distinct parasite clade. *Int. J. Parasitol. Parasites Wildl.* **7**, 207–212 (2018).

77. Ogedengbe, J. D., Ogedengbe, M. E., Hafeez, M. A. & Barta, J. R. Molecular phylogenetics of eimeriid coccidia (Eimeriidae, Eimeriorina, Apicomplexa, Alveolata): A preliminary multi-gene and multi-genome approach. *Parasitol. Res.* **114**, 4149–4160 (2015).

78. Ogedengbe, M. E., El-Sherry, S., Ogedengbe, J. D., Chapman, H. D. & Barta, J. R. Phylogenies based on combined mitochondrial and nuclear sequences conflict with morphologically defined genera in the eimeriid coccidia (Apicomplexa). *Int. J. Parasitol.* **48**, 59–69 (2018).

79. Dupuis, J. R., Roe, A. D. & Sperling, F. A. H. Multi-locus species delimitation in closely related animals and fungi: One marker is not enough. *Mol. Ecol.* **21**, 4422–4436 (2012).

80. Marcelino, V. R. & Verbruggen, H. Multi-marker metabarcoding of coral skeletons reveals a rich microbiome and diverse evolutionary origins of endolithic algae. *Sci. Rep.* **6**, 1–9 (2016).

81. Mallo, D. & Posada, D. Multilocus inference of species trees and DNA barcoding. *Philos. Trans. R. Soc. B Biol. Sci.* **371**, (2016).

82. Gevers, D. *et al.* Re-evaluating prokaryotic species. *Nat. Rev. Microbiol.* **3**, 733–739 (2005).

83. Glaeser, S. P. & Kämpfer, P. Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Syst. Appl. Microbiol.* **38**, 237–245 (2015).

84. Feng, Y. *et al.* Development of a multilocus sequence typing tool for high-resolution genotyping of *Enterocytozoon bieneusi*. *Appl. Environ. Microbiol.* **77**, 4822–4828 (2011).

85. Boité, M. C., Mauricio, I. L., Miles, M. A. & Cupolillo, E. New Insights on Taxonomy, Phylogeny and Population Genetics of *Leishmania* (*Viannia*) Parasites Based on Multilocus Sequence Analysis. *PLoS Negl. Trop. Dis.* **6**, (2012).

86. Lauthier, J. J. *et al.* Candidate targets for Multilocus Sequence Typing of *Trypanosoma cruzi*: Validation using parasite stocks from the Chaco Region and a set of reference strains. *Infect. Genet. Evol.* **12**, 350–358 (2012).

87. Lefoulon, E. *et al.* Shaking the Tree: Multi-locus Sequence Typing Usurps Current Onchocercid (Filarial Nematode) Phylogeny. *PLoS Negl. Trop. Dis.* **9**, 1–19 (2015).

88. Guo, Y. *et al.* Multilocus sequence typing tool for *Cyclospora cayetanensis*. *Emerg. Infect. Dis.* **22**, 1464–1467 (2016).

89. Tenter, A. M. *et al.* The conceptual basis for a new classification of the coccidia. *Int. J. Parasitol.* **32**, 595–616 (2002).

90. Kvičerová, J. & Hypša, V. Host-Parasite Incongruences in Rodent *Eimeria* Suggest Significant Role of Adaptation Rather than Cophylogeny in Maintenance of Host Specificity. *PLoS One* **8**, (2013).

91. Mácová, A. *et al.* Mysteries of host switching: Diversification and host specificity in rodent-coccidia associations. *Mol. Phylogenet. Evol.* **127**, 179–189 (2018).

92. Leung, T. L. F. & Poulin, R. Parasitism, commensalism, mutualism: Exploring the many shades of symbioses. *Vie Milieu - Life Environ.* **58**, 107–115 (2008).

93. Leggett, H. C., Brown, S. P. & Reece, S. E. War and peace: social interactions in infections Subject collections War and peace: social interactions in infections. *Philos. Trans. R. Soc. B Biol. Sci.* **369**, (2013).

94. Poulin, R. Host specificity. in *Evolutionary Ecology of Parasites* 42–69 (Princeton University Press, 2007).

95. Schmid-Hempel, P. *Evolutionary Parasitology the integrated study of infections, immunology, ecology, and genetics.* (Oxford University Press, 2011).

96. Lymbery, A. J. Host specificity, host range and host preference. *Parasitol. Today* **5**, 298 (1989).

97. Dallas, T., Huang, S., Nunn, C., Park, A. W. & Drake, J. M. Estimating parasite host

range. *Proc. R. Soc. B Biol. Sci.* **284**, (2017).

98. Combes, C. *Parasitism: the ecology and evolution of intimate interactions.* (University of Chicago Press, 2001).

99. Rigaud, T., Perrot-Minnot, M. J. & Brown, M. J. F. Parasite and host assemblages: Embracing the reality will improve our knowledge of parasite transmission and virulence. *Proc. R. Soc. B Biol. Sci.* **277**, 3693–3702 (2010).

100. Poulin, R. & Mouillot, D. Combining phylogenetic and ecological information into a new index of host specificity. *J. Parasitol.* **91**, 511–514 (2005).

101. Poulin, R. & Rohde, K. Comparing the richness of metazoan ectoparasite communities of marine fishes: Controlling for host phylogeny. *Oecologia* **110**, 278–283 (1997).

102. Dick, C. W. High host specificity of obligate ectoparasites. *Ecol. Entomol.* **32**, 446–450 (2007).

103. Poulin, R. Parasite Population Dynamics and Genetics. in *Evolutionary Ecology of Parasites* 160–187 (Princeton University Press, 2007).

104. Ebert, D. Transmitted Parasite. *Science (80-. ).* **265**, 16–18 (1994).

105. Kaltz, O. & Shykoff, J. A. Local adaptation in host-parasite systems. *Heredity (Edinb).* **81**, 361–370 (1998).

106. Kaci-Chaouch, T., Verneau, O. & Desdevises, Y. Host specificity is linked to intraspecific variability in the genus *Lamellodiscus* (Monogenea). *Parasitology* **135**, 607–616 (2008).

107. Poulin, R. & Keeney, D. B. Host specificity under molecular and experimental scrutiny. *Trends Parasitol.* **24**, 24–28 (2008).

108. Becker, E. R. *Coccidia and Coccidiosis of Domesticated, Game and Laboratory Animals and of Man*. *JAMA: The Journal of the American Medical Association* **102**, (Akademiai Kiado., 1934).

109. Kheysin, Y. M. *Life cycles of Coccidia of domestic animals*. (William Heineman Medical Books Limited, 1972).

110. Pellérdy, L. P. *Coccidia and coccidiosis*. (Parey, 1974).

111. Berman, J. J. Apicomplexa. in *Taxonomic Guide to Infectious Diseases* 103–109 (Elsevier, 2012).

112. Levine, N. D. & Ivens, V. *The Coccidian Parasites (Protozoa, Sporozoa) of Rodents*. (The University of Ilinois Press, 1965).

113. Levine, N. D. & Ivens, V. *The coccidian parasites (Protozoa, Sporozoa) of ruminants*. (University of Illinois Press, 1970).

114. Davis, L. R., Levine, N. D. & Ivens, V. The Coccidian Parasites (Protozoa, Sporozoa) of Ruminants. *J. Parasitol.* **57**, 558 (1971).

115. Levine, N. D. & Ivens, V. *The coccidian parasites (Protozoa, Apicomplexa) of carnivores*. (University of Illinois Press, 1981).

116. Dyková, I. & Lom, J. Fish coccidia: critical notes on life cycles, classification and pathogenicity. *J. Fish Dis.* **4**, 487–505 (1981).

117. Duszynski, D. W. *Eimeria*. in *eLS* **95**, 1192–1196 (John Wiley & Sons, Ltd, 2001).

118. Marquardt, W. C. Host and site specifity in the coccidia. **Univ. Park**, (1973).

119. Smith, A. L. & Hayday, A. C. Genetic dissection of primary and secondary responses to a widespread natural pathogen of the gut, *Eimeria vermiformis*. *Infect. Immun.* **68**, 6273–6280 (2000).

120. Wang, S. & Suo, X. Still naïve or primed: Anticoccidial vaccines call for memory. *Exp. Parasitol.* **216**, 107945 (2020).

121. Dalloul, R. A. & Lillehoj, H. S. Poultry coccidiosis: Recent advancements in control measures and vaccine development. *Expert Rev. Vaccines* **5**, 143–163 (2006).

122. Blake, D. P. & Tomley, F. M. Securing poultry production from the ever-present *Eimeria* challenge. *Trends Parasitol.* **30**, 12–19 (2014).

123. Berto, B. P., Mcintosh, D. & Lopes, C. W. G. Studies on coccidian oocysts (Apicomplexa: Eucoccidiorida). *Rev. Bras. Parasitol. Vet.* **23**, 1–15 (2014).

124. Long, P. L. & Joyner, L. P. Problems in the identification of species of *Eimeria*. *J. Protozool.* **31**, 535–541 (1984).

125. Chapman, H. D. *et al. A selective review of advances in coccidiosis research*. *Advances in Parasitology* **83**, (Elsevier Ltd., 2013).

126. Parker, B. B. & Duszynski, D. W. Polymorphism of Eimerian Oocysts : A Dilemma Posed by Working with Some Naturally Infected Hosts. *J. Parasitol.* **72**, 602–604 (1986).

127. Gardner, S. L. & Duszynski, D. W. Polymorphism of Eimerian oocysts can be a problem in naturally infected hosts: An example from subterranean rodents in Bolivia. *J. Parasitol.* **76**, 805–811 (1990).

128. Fayer, R. Epidemiology of protozoan infections: The coccidia. *Vet. Parasitol.* **6**, 75–103 (1980).

129. Duszynski, D. W. & Wilber, P. G. A guidelines for the preparation of species description in the Eimeriidae. *J. Parasitol.* **83**, 333–336 (1997).

130. Hnida, J. A. & Duszynski, D. W. Taxonomy and systematics of some *Eimeria* species of murid rodents as determined by the ITS1 region of the ribosomal gene complex. *Parasitology* **119**, 349–357 (1999).

131. Zhao, X. & Duszynski, D. Molecular phylogenies suggest the oocyst residuum can be used to distinguish two independent lineages of *Eimeria* spp in rodents. *Parasitol. Res.* **87**, 638–643 (2001).

132. Kvičerová, J., Pakandl, M. & Hypša, V. Phylogenetic relationships among *Eimeria* spp. (Apicomplexa, Eimeriidae) infecting rabbits: Evolutionary significance of biological and morphological features. *Parasitology* **135**, 443–452 (2008).

133. Duszynski, D. W. & Upton, S. J. *Cyclospora, Eimeria, Isospora* and *Cryptosporidium* spp. in *Parasitic Diseases of Wild Mammals* (eds. Samuel, W. M., Pybus, M. J. & Kocan, A. A.) 416–459 (Iowa State University Press, 2001).

134. Kay, E. H. & Hoekstra, H. E. Rodents. *Curr. Biol.* **18**, 406–410 (2008).

135. Fuller, C. A. & Blaustein, A. R. Effects of the parasite Eimeria arizonensis on survival of deer mice (Peromyscus maniculatus). *Ecology* **77**, 2196–2202 (1996).

136. Hnida, J. A., Wilson, W. D. & Duszynski, D. W. *Onychomys* Infecting a New *Eimeria* Species (Apicomplexa : Eimeriidae) and Arizona in New Mexico Species (Rodentia : Muridae). *J. Parasitol* **84**, 1207–1209 (1998).

137. Kvičerová, J., Mikeš, V. & Hypša, V. Third lineage of rodent eimerians: Morphology, phylogeny and re-description of *Eimeria myoxi* (Apicomplexa: Eimeriidae) from Eliomys quercinus (Rodentia: Gliridae). *Parasitology* **138**, 1217–1223 (2011).

138. Eimer, T. *Ueber die ei-oder kugelförmigen sogenannten Psorospermien der Wirbelthiere: Ein Beitrag Zur Entwicklungsgeschichte Der Gregarinen Und Zur Kenntniss Dieser Parasiten Als Krankheitsursache.* (A. Stuber., 1870).

139. Schneider, A. Note sur la psorospermie oviforme du poulpe. *Arch. Zool. exp. gen,* **4**, 4045 (1875).

140. Schmid, M., Lehmann, M. J., Lucius, R. & Gupta, N. Apicomplexan parasite, *Eimeria falciformis*, co-opts host tryptophan catabolism for life cycle progression in mouse. *J. Biol. Chem.* **287**, 20197–20207 (2012).

141. Ehret, T., Spork, S., Dieterich, C., Lucius, R. & Heitlinger, E. Dual RNA-seq reveals no plastic transcriptional response of the coccidian parasite *Eimeria falciformis* to host immune defenses. *BMC Genomics* **18**, 1–17 (2017).

142. Huang, G. *et al.* Influence of *Eimeria falciformis* infection on gut microbiota and metabolic pathways in mice. *Infect. Immun.* **86**, (2018).

143. Heitlinger, E., Spork, S., Lucius, R. & Dieterich, C. The genome of *Eimeria falciformis* - reduction and specialization in a single host apicomplexan parasite. *BMC Genomics* **15**, 696 (2014).

144. Al-khlifeh, E. *et al. Eimeria falciformis* BayerHaberkorn1970 and novel wild derived isolates from house mice : differences in parasite lifecycle, pathogenicity and host immune reactions. *bioRxiv* 1–33 (2019). doi:http://dx.doi.org/10.1101/611277

145. Balard, A. *et al.* Coupling between tolerance and resistance differs between related *Eimeria* parasite species: implications for coevolution with their mouse hosts. *bioRxiv* 2020.01.24.918144 (2020). doi:10.1101/2020.01.24.918144

146. de Vos, A. J. Studies on the host range of *Eimeria chinchillae* De Vos & Van der Westhuizen, 1968. *Onderstepoort J. Vet. Res.* **37**, 29–36 (1970).

147. Levine, N. D. & Ivens, V. Cross-Transmission of *Eimeria* spp. (Protozoa, Apicomplexa) of Rodents -A Review. *J. Protozool.* **35**, 434–437 (1988).

148. Nowell, F. & Higgs, S. *Eimeria* species infecting wood mice (genus *Apodemus*) and the transfer of two species to *Mus musculus*. *Parasitology* **98**, 329–336 (1989).

149. Kvičerová, J., Ptáčková, P. & Modrý, D. Endogenous development, pathogenicity and host specificity of *Eimeria cahirinensis* Couch, Blaustein, Duszynski, Shenbrot and Nevo, 1997 (Apicomplexa: Eimeriidae) from *Acomys dimidiatus* (Cretzschmar 1826) (Rodentia: Muridae) from the Near Eas. *Parasitol. Res.* **100**, 219–226 (2007).

150. Kvicerova, J., Hofmannova, L., Scognamiglio, F. & Santoro, M. *Eimeria sciurorum* (Apicomplexa, Coccidia) From the Calabrian Black Squirrel (*Sciurus meridionalis*): An Example of Lower Host Specificity of Eimerians. *Front. Vet. Sci.* **7**, 6–8 (2020).

151. Houdebine, L.-M. *Chapter 6 - The Mouse as an Animal Model for Human Diseases BT - The laboratory mouse*. *The laboratory mouse* (Elsevier Ltd, 2004). doi:10.1016/B978-0-12-336425-8.50059-5

152. Vandenbergh, J. G. Use of House Mice in Biomedical Research. *ILAR J.* **41**, 133–135 (2013).

153. Nishioka, Y. The origin of common laboratory mice. *Genome* **38**, 1–7 (2011).

154. Yang, H., Bell, T. A., Churchill, G. A. & Pardo-Manuel De Villena, F. On the subspecific origin of the laboratory mouse. *Nat. Genet.* **39**, 1100–1107 (2007).

155. Ehret, T., Torelli, F., Klotz, C., Pedersen, A. B. & Seeber, F. Translational Rodent Models for Research on Parasitic Protozoa—A Review of Confounders and Possibilities. *Front. Cell. Infect. Microbiol.* **7**, (2017).

156. Perkins, F. O., Barta, J. R., Clopton, R. E., Peirce, M. A. & Upton, S. J. Apicomplexa. in *The Illustrated Guide to the Protozoa* (eds. Lee, J., Leedale, G. & Bradbury, P.) 190–304 (Allen Press, 2000).

157. Shirley, M. W., Smith, A. L. & Tomley, F. M. The biology of avian *Eimeria* with an emphasis on their control by vaccination. *Adv. Parasitol.* **60**, 285–330 (2005).

158. Su, Y. C., Fei, A. C. Y. & Tsai, F. M. Differential diagnosis of five avian Eimeria species by polymerase chain reaction using primers derived from the internal transcribed spacer 1 (ITS-1) sequence. *Vet. Parasitol.* **117**, 221–227 (2003).

159. Zhao, X. & Duszynski, D. W. Phylogenetic relationships among rodent *Eimeria* species determined by plastid ORF470 and nuclear 18S rDNA sequences. *Int. J. Parasitol.* **31**, 715–719 (2001).

160. Ryley, J. F., Meade, R., Hazelhurst, J. & Robinson, T. E. Methods in coccidiosis research: Separation of oocysts from faeces. *Parasitology* **73**, 311–326 (1976).

161. Fernandez, S. *et al.* A multiplex PCR assay for the simultaneous detection and discrimination of the seven *Eimeria* species that infect domestic fowl. *Parasitology* **127**, 317–325 (2003).

162. Hnida, J. A. & Duszynski, D. W. Taxonomy and phylogeny of some *Eimeria* (Apicomplexa: Eimeriidae) species of rodents as determined by polymerase chain

reaction/restriction-fragment-length polymorphism analysis of 18S rDNA. *Parasitol. Res.* **85**, 887–94 (1999).

163. Kawahara, F. *et al.* Genetic analysis and development of species-specific PCR assays based on ITS-1 region of rRNA in bovine *Eimeria* parasites. *Vet. Parasitol.* **174**, 49–57 (2010).

164. Morris, G. M., Woods, W. G., Grant Richards, D. & Gasser, R. B. The application of a polymerase chain reaction (PCR)-based capillary electrophoretic technique provides detailed insights into *Eimeria* populations in intensive poultry establishments. *Mol. Cell. Probes* **21**, 288–294 (2007).

165. Schnitzler, B. E., Thebo, P. L., Tomley, F. M., Uggla, A. & Shirley, M. W. PCR identification of chicken *Eimeria*: A simplified read-out. *Avian Pathol.* **28**, 89–93 (1999).

166. Vrba, V., Blake, D. P. & Poplstein, M. Quantitative real-time PCR assays for detection and quantification of all seven *Eimeria* species that infect the chicken. *Vet. Parasitol.* **174**, 183–190 (2010).

167. Owen, D. Some parasites and other organisms of wild rodents in the vicinity of an SPF unit. *Lab. Anim.* **10**, 271–278 (1976).

168. Haberkorn, A. Die Entwicklung von Eimeria falciformis (Eimer 1870) in der weissen Maus (Mus musculus). *Zeitschrift für Parasitenkunder* **34**, 49–67 (1970).

169. Mesfin, G. M. & Bellamy, J. E. C. The Life Cycle of *Eimeria falciformis* var. *Pragensis* (Sporozoa: Coccidia) in the Mouse, *Mus musculus*. *J. Parasitol.* **64**, 696–705 (1978).

170. Schelzke, K. *et al.* CD8+ cells protect mice against reinfection with the intestinal parasite *Eimeria falciformis*. *Microbes Infect.* **12**, 218–226 (2009).

171. Ernst, J. V., Chobotar, B. & Hammond, D. M. The Oocysts of *Eimeria vermiformis* sp. n. and *E. papillata* sp. n. (Protozoa: Eimeriidae) from the Mouse *Mus musculus*. *J. Protozool.* **18**, 221–223 (1971).

172. Figueiredo-Campos, P., Ferreira, C., Blankenhaus, B. & Veldhoen, M. *Eimeria vermiformis* Infection Model of Murine Small Intestine. *BIO-PROTOCOL* **8**, (2018).

173. Rose, M. E., Wakelin, D. & Hesketh, P. *Eimeria vermiformis*: Differences in the course of primary infection can be correlated with lymphocyte responsiveness in the BALB/c and C57BL/6 mouse, *Mus musculus*. *Exp. Parasitol.* **71**, 276–283 (1990).

174. Rose, M. E. & Millard, B. J. *Eimeria vermiformis*: Host strains and the developmental cycle. *Exp. Parasitol.* **60**, 285–293 (1985).

175. Todd Jr, K. T. & Lepp, D. L. The Life Cycle of *Eimeria vermiformis* Ernst, Chobotar and Hammond, 1971 in the Mouse *Mus musculus*. *J. Protozool.* **18**, 332–337 (1971).

176. Ankrom, S. L., Chobotar, B. & Ernest, J. V. Life Cycle of *Eimeria ferrisi* Levine & Ivens, 1965 in the Mouse, *Mus musculus*. *J. Protozool.* **22**, 317–323 (1975).

177. Schito, M. L., Barta, J. R. & Chobotar, B. Comparison of four murine *Eimeria* species in immunocompetent and immunodeficient mice. *J. Parasitol.* **82**, 255–262 (1996).

178. Ball, S. J. & Lewis, D. C. *Eimeria* (Protozoa: Coccidia) in wild populations of some British rodents. *J. Zool.* **202**, 373–381 (1984).

179. Golemanski, V. G. The coccidia (Coccidia, Eimeriidae) of small mammals from the Parangalitsa, Ropotamo and Srebirna Reserves in Bulgaria. *Acta Zool. Bulg.* **12**, 12–26 (1979).

180. Owen D. Some parasites and other organisms of wild rodents in the vicinity of an SPF unit. *Lab. Anim.* **10**, 271–278 (1976).

181. Parker, S. E., Malone, S., Bunte, R. M. & Smith, A. L. Infectious diseases in wild mice (Mus musculus) collected on and around the University of Pennsylvania (Philadelphia) Campus. *Comp. Med.* **59**, 424–430 (2009).

182. Yakimoff, W. L. & Gousseff, W. F. The coccidia of mice Mus musculus. *Parasitology* **30**, 26–28 (1938).

183. Boursot, P., Auffray, J.-C., Britton-Davidian, J. & Bonhomme, F. The Evolution of House Mice. *Annu. Rev. Ecol. Syst.* **24**, 119–152 (2003).

184. Ďureje, L., Macholán, M., Baird, S. J. E. & Piálek, J. The mouse hybrid zone in Central Europe: From morphology to molecules. *Folia Zool.* **61**, 308–318 (2012).

185. Phifer-Rixey, M. & Nachman, M. W. Insights into mammalian biology from the wild house mouse *Mus musculus*. *Elife* **2015**, 1–13 (2015).

186. Ahmed, N. *et al.* A Novel Non-invasive Method to Detect RELM Beta Transcript in Gut Barrier Related Changes During a Gastrointestinal Nematode Infection. *Front. Immunol.* **10**, 445 (2019).

187. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

188. Guindon, S. *et al.* New Algorithms and Mehtods to Estimate Maximum-Likelihood Phylogenies: Asessing the Performance of PhyML 2.0. *Syst. Biol.* **59**, 307–321 (2010).

189. Huelsenbeck, J. P. & Ronquist, F. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).

190. Ronquist, F. *et al.* Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).

191. Posada, D. jModelTest: Phylogenetic model averaging. *Mol. Biol. Evol.* **25**, 1253–1256 (2008).

192. Rambaut, A. FigTree. (2012).

193. Sterne, T. E. Some remarks on confidence or fiducial limits. *Biometrika* **41**, 275–278 (1954).

194. Nunes, M. S. *et al.* epiR: Tools for the analysis of epidemiological data. (2018).

195. Fisher, R. A. On the Interpretation of χ2 from Contingency Tables, and the Calculation of P. *J. R. Stat. Soc.* **85**, 87 (1922).

196. Long, J. jtools: Analysis and Presentation of Social Scientific Data. (2019).

197. Tattersall, F. H., Nowell, F. & Smith, R. H. A review of the endoparasites of wild House Mice *Mus domesticus*. *Mamm. Rev.* **24**, 61–71 (1994).

198. Moro, D., Lawson, M. A., Hobbs, R. P. & Thompson, R. C. A. Pathogens of House Mice on Arid Boullanger Island and Subantarctic Macquarie Island, Australia. *J. Wildl. Dis.* **39**, 762–771 (2003).

199. Hobbs, R. P., Twigg, L. E., Elliot, A. D. & Wheeler, A. G. Factors Influencing the Fecal Egg and Oocyst Counts of Parasites of Wild European Rabbits *Oryctolagus cuniculus* (L.) in Southern Western Australia. *J. Parasitol.* **85**, 796–802 (1999).

200. Hu, X. L. *et al.* Methods of preservation and flotation for the detection of nematode eggs and coccidian oocysts in faeces of the forest musk deer. *J. Helminthol.* **90**, 680–684 (2016).

201. Rinaldi, L., Coles, G. C., Maurelli, M. P., Musella, V. & Cringoli, G. Calibration and diagnostic accuracy of simple flotation, McMaster and FLOTAC for parasite egg counts in sheep. *Vet. Parasitol.* **177**, 345–352 (2011).

202. Ballweber, L. R., Beugnet, F., Marchiondo, A. A. & Payne, P. A. American Association of Veterinary Parasitologists' review of veterinary fecal flotation methods and factors influencing their accuracy and use—Is there really one best technique? *Vet. Parasitol.* **204**, 73–80 (2014).

203. Webster, K. A., Smith, H. V., Giles, M., Dawson, L. & Robertson, L. J. Detection of *Cryptosporidium parvum* oocysts in faeces: Comparison of conventional coproscopical methods and the polymerase chain reaction. *Vet. Parasitol.* **61**, 5–13 (1996).

204. Raj, G. D. *et al.* Real-time PCR-based quantification of *Eimeria* genomes: A method to outweigh underestimation of genome numbers due to PCR inhibition. *Avian Pathol.* **42**, 304–308 (2013).

205. Yang, R., Brice, B. & Ryan, U. Morphological and molecular characterization of *Eimeria haematodi*, coccidian parasite (Apicomplexa: Eimeriidae) in a rainbow lorikeet (*Trichoglossus haematodus*). *Exp. Parasitol.* **153**, 123–128 (2015).

206. Šlapeta, J. R. *et al.* Eimeria telekii n.sp. (Apicomplexa: Coccidia) from *Lemniscomys striatus* (Rodentia: Muridae): Morphology, pathology and phylogeny. *Parasitology* **122**, 133–143 (2001).

207. Higgs, S. & Nowell, F. Population biology of *Eimeria* (Protozoa: Apicomplexa) in *Apodemus sylvaticus*: A capture/recapture study. *Parasitology* **120**, 355–363 (2000).

208. Hofmannová, L. *et al.* Diversity and host specificity of coccidia (Apicomplexa: Eimeriidae) in native and introduced squirrel species. *Eur. J. Protistol.* **56**, 1–14 (2016).

209. Goüy de Bellocq, J. *et al.* Holobiont suture zones: Parasite evidence across the European house mouse hybrid zone. *Mol. Ecol.* **27**, 5214–5227 (2018).

210. Kváč, M. *et al.* Coevolution of *Cryptosporidium tyzzeri* and the house mouse (*Mus*

*musculus*). *Int. J. Parasitol.* **43**, 805–817 (2013).

211. Tompkins, D. M., Dunn, A. M., Smith, M. J. & Telfer, S. Wildlife diseases: From individuals to ecosystems. *J. Anim. Ecol.* **80**, 19–38 (2011).

212. Adler, F. R., Pearce-Duvet, J. M. C. & Dearing, M. D. How host population dynamics translate into time-lagged prevalence: An Investigation of Sin Nombre virus in deer mice. *Bull. Math. Biol.* **70**, 236–252 (2008).

213. Khalil, H. *et al.* Dynamics and Drivers of Hantavirus Prevalence in Rodent Populations. *Vector-Borne Zoonotic Dis.* **14**, 537–551 (2014).

214. Sauvage, F., Langlais, M., Yoccoz, N. G. & Pontier, D. Modelling hantavirus on virus persistence voles : the role of indirect transmission. *J. Anim. Ecol.* **72**, 1–13 (2003).

215. Singleton, G. R., Smith, A. L. & Krebs, C. J. The prevalence of viral antibodies during a large population fluctuation of house mice in Australia. *Epidemiol. Infect.* **125**, 719–727 (2000).

216. Haukisalmi, V. & Henttonen, H. The impact of climatic factors and host density on the long-term population dynamics of vole helminths. *Oecologia* **83**, 309–315 (1990).

217. Scott, M. E. & Lewis, J. W. Population dynamics of helminth parasites in wild and laboratory rodents. *Mamm. Rev.* **17**, 95–103 (1987).

218. Anderson, R. M. & May, R. M. Coevolution of Hosts and Parasites. *Parasitology* **85**, 411–426 (1982).

219. Frank, S. A. Models of Parasite Virulence. *Q. Rev. Biol.* **71**, 37–78 (1996).

220. Tilahun, G. & Stockdale, P. H. G. Oocyst production of four species of murine coccidia. *Can. J. Zool.* **59**, 1796–1800 (1981).

221. Marquardt, W. C. Host and site specifity in the coccidia. *J. Protozool.* **28**, 243–244 (1981).

222. Wilber, P. G., Duszynski, D. W., Upton, S. J., Seville, R. S. & Corliss, J. O. A revision of the taxonomy and nomenclature of the *Eimeria* spp. (Apicomplexa: Eimeriidae) from rodents in the Tribe Marmotini (Sciuridae). *Syst. Parasitol.* **39**, 113–135 (1998).

223. Upton, S. J., McAllister, C. T., Brillhart, D. B., Duszynski, D. W. & Wash, C. D. Cross-Transmission Studies with *Eimeria arizonensis*-like Oocysts (Apicomplexa) in New World Rodents of the Genera *Baiomys, Neotoma, Onychomys, Peromyscus*, and *Reithrodontomys* (Muridae). *J. Parasitol.* **78**, 406–413 (1992).

224. Duszynski, D. W., Eastham, G. & Yates, T. L. *Eimeria* from Jumping Mice (*Zapus* spp.): A New Species and Genetic and Geographic Features of *Z. hudsonius luteus*. *J. Parasitol.* **68**, 1146–1148 (1982).

225. Lainson, R. & Shaw, J. J. Coccidia of Brazilian Mammals: *Eimeria corticulata* N. Sp. (Apicomplexa: Eimeriidae) from the Anteater *Tamandua tetradactyla* (Xenarthra: Myrmecophagidae) and *Eimeria zygodontomyis* N. Sp. from the Cane Mouse *Zygodontomys lasiurus*. *J. Protozool.* **37**, 51–54 (1990).

226. Mesfin, G. M., Bellamy, J. E. & Stockdale, P. H. The pathological changes caused by *Eimeria falciformis* var pragensis in mice. *Can. J. Comp. Med. Rev. Can. Med. Comp.* **42**, 496–510 (1978).

227. Todd, K. S. & Hammond, D. M. Life Cycle and Host Specificity of *Eimeria callospermophili* Henry, 1932 From the Uinta Ground Squirrel *Spermophilus armatus*. *J. Protozool.* **15**, 1–8 (1968).

228. Turner, W. C., Penzhorn, B. L. & Getz, W. M. Description of 3 New Species of *Eimeria* (Apicomplexa: Eimeriidae) from Springbok (*Antidorcas marsupialis*) in Namibia. *Comp. Parasitol.* **83**, 202–211 (2016).

229. Wash, C. D., Duszynski, D. W. & Yates, T. L. Eimerians from Different Karyotypes of the Japanese Wood Mouse (*Apodemus* spp.), with Descriptions of Two New Species and a Redescription of *Eimeria montgomeryae* Lewis and Ball, 1983. *J. Parasitol.* **71**, 808–814 (1985).

230. MacPherson, J. M. & Gajadhar, A. A. Differentiation of seven *Eimeria* species by random amplified polymorphic DNA. *Vet. Parasitol.* **45**, 257–266 (1993).

231. Barta, J. R. *et al.* Phylogenetic Relationships among Eight *Eimeria* Species Infecting Domestic Fowl Inferred Using Complete Small Subunit Ribosomal DNA Sequences. *J. Parasitol.* **83**, 262 (1997).

232. Pellérdy, L. Zur Kenntnis der Coccidien aus *Apodemus flavicollis*. *Acta Vet. Acad. Sci. Hungaricae* **4**, 187–191 (1954).

233. Jarquín-Díaz, V. H. *et al.* Detection and quantification of house mouse *Eimeria* at the species level – Challenges and solutions for the assessment of coccidia in wildlife. *Int. J. Parasitol. Parasites Wildl.* **10**, 29–40 (2019).

234. Reutter, B. A., Petit, E., Brünner, H. & Vogel, P. Cytochrome b haptotype divergences in West European *Apodemus*. *Mamm. Biol.* **68**, 153–164 (2003).

235. Abramson, N. I., Rodchenkova, E. N. & Kostygov, A. Y. Genetic variation and phylogeography of the bank vole (*Clethrionomys glareolus*, Arvicolinae, Rodentia) in Russia with special reference to the introgression of the mtDNA of a closely related species, red-backed vole (*Cl. rutilus*). *Russ. J. Genet.* **45**, 533–545 (2009).

236. Jaarola, M. & Searle, J. B. Phylogeography of field voles (*Microtus agrestis*) in Eurasia inferred from mitochondrial DNA sequences. *Mol. Ecol.* **11**, 2613–2621 (2002).

237. Paradis, E., Thibaut, J., Knaus, B., Klaus, S. & Alastair, P. Package 'pegas' Population and Evolutionary Genetics Analysis System. (2018).

238. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).

239. Heitlinger, E. MultiAmplicon. (2019).

240. Wright, E. S. Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R. *R J.* **8**, 352–359 (2016).

241. Dray, S. & Dufour, A.-B. The ade4 Package: Implementing the Duality Diagram for Ecologists. *J. Stat. Softw.* **22**, (2015).

242. Jombart, T. Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).

243. Deng, W. *et al.* DIVEIN: A web server to analyze phylogenies, sequence divergence, diversity, and informative sites. *Biotechniques* **48**, 405–408 (2010).

244. Sukumaran, J. & Holder, M. T. DendroPy: A Python library for phylogenetic computing. *Bioinformatics* **26**, 1569–1571 (2010).

245. Reid, A. J. *et al.* Genomic analysis of the causative agents of coccidiosis in domestic chickens. *Genome Res.* **24**, 1676–1685 (2014).

246. Meyer, C. P. & Paulay, G. DNA barcoding: Error rates based on comprehensive sampling. *PLoS Biol.* **3**, 1–10 (2005).

247. Churakov, G. *et al.* Rodent evolution: Back to the root. *Mol. Biol. Evol.* **27**, 1315–1326 (2010).

248. Steppan, S. J., Adkins, R. M. & Anderson, J. Phylogeny and divergence-date estimates of rapid radiations in muroid rodents based on multiple nuclear genes. *Syst. Biol.* **53**, 533–553 (2004).

249. Adamson, M. L. & Caira, J. N. Evolutionary factors influencing the nature of parasite specificity. *Parasitology* **109**, S85–S95 (1994).

250. Poulin, R., Krasnov, B. R. & Mouillot, D. Host specificity in phylogenetic and geographic space. *Trends Parasitol.* **27**, 355–361 (2011).

251. Fenton, A. & Brockhurst, M. A. The role of specialist parasites in structuring host communities. *Ecol. Res.* **23**, 795–804 (2008).

252. Forbes, M. R., Muma, K. E. & Smith, B. P. Diffuse coevolution: Constraints on a generalist parasite favor use of a dead-end host. *Ecography (Cop.).* **25**, 345–351 (2002).

253. Kassen, R. The experimental evolution of specialists, generalist, and the maintenance of diversity. *J. Evol. Biol.* **15**, 173–190 (2002).

254. Rathore, D. *et al.* Molecular Mechanism of Host Specificity in *Plasmodium falciparum* Infection. *J. Biol. Chem.* **278**, 40905–40910 (2003).

255. Carreno, R. A., Martin, D. S. & Barta, J. R. Cryptosporidium is more closely related to the gregarines than to coccidia as shown by phylogenetic analysis of apicomplexan parasites inferred using small-subunit ribosomal RNA gene sequences. *Parasitol. Res.* **85**, 899–904 (1999).

256. Barta, J. R. Molecular approaches for inferring evolutionary relationships among protistan parasites. *Vet. Parasitol.* **101**, 175–186 (2001).

257. Vrba, V., Poplstein, M. & Pakandl, M. The discovery of the two types of small subunit ribosomal RNA gene in *Eimeria mitis* contests the existence of *E. mivati* as an

independent species. *Vet. Parasitol.* **183**, 47–53 (2011).

258. El-Sherry, S., Ogedengbe, M. E., Hafeez, M. A. & Barta, J. R. Divergent nuclear 18S rDNA paralogs in a turkey coccidium, *Eimeria meleagrimitis*, complicate molecular systematics and identification. *Int. J. Parasitol.* **43**, 679–685 (2013).

259. Baum, D. A. & Donoghue, M. J. Choosing among alternative 'Phylogenetic' Species Concept. *Syst. Bot.* **20**, 560–573 (1995).

260. Lewis, D. C. & Ball, S. J. Species of *Eimeria* of small wild rodents from the British Isles, with descriptions of two new species. *Syst. Parasitol.* **5**, 259–270 (1983).

261. Higgs, S. & Nowell, F. A review of the species of Eimeria infecting hosts in the genus *Apodemus*. *Syst. Parasitol.* **20**, 203–209 (1991).

262. Seville, R. S. & Stanton, N. L. Synonymy of *Eimeria larimerensis* with *Eimeria lateralis*. *J. Parasitol.* **79**, 970–972 (1993).

263. El-Sherry, S. *et al.* Sequence-based genotyping clarifies conflicting historical morphometric and biological data for 5 *Eimeria* species infecting Turkeys. *Poult. Sci.* **94**, 262–272 (2015).

# Appendix (Supplementary material)

The following supplementary materials from **Chapter 2** correspond to tables containing raw data not included in the present thesis:

**Supplementary data S2.1 Geographical coordinates for collected mice in Brandenburg, Germany.**

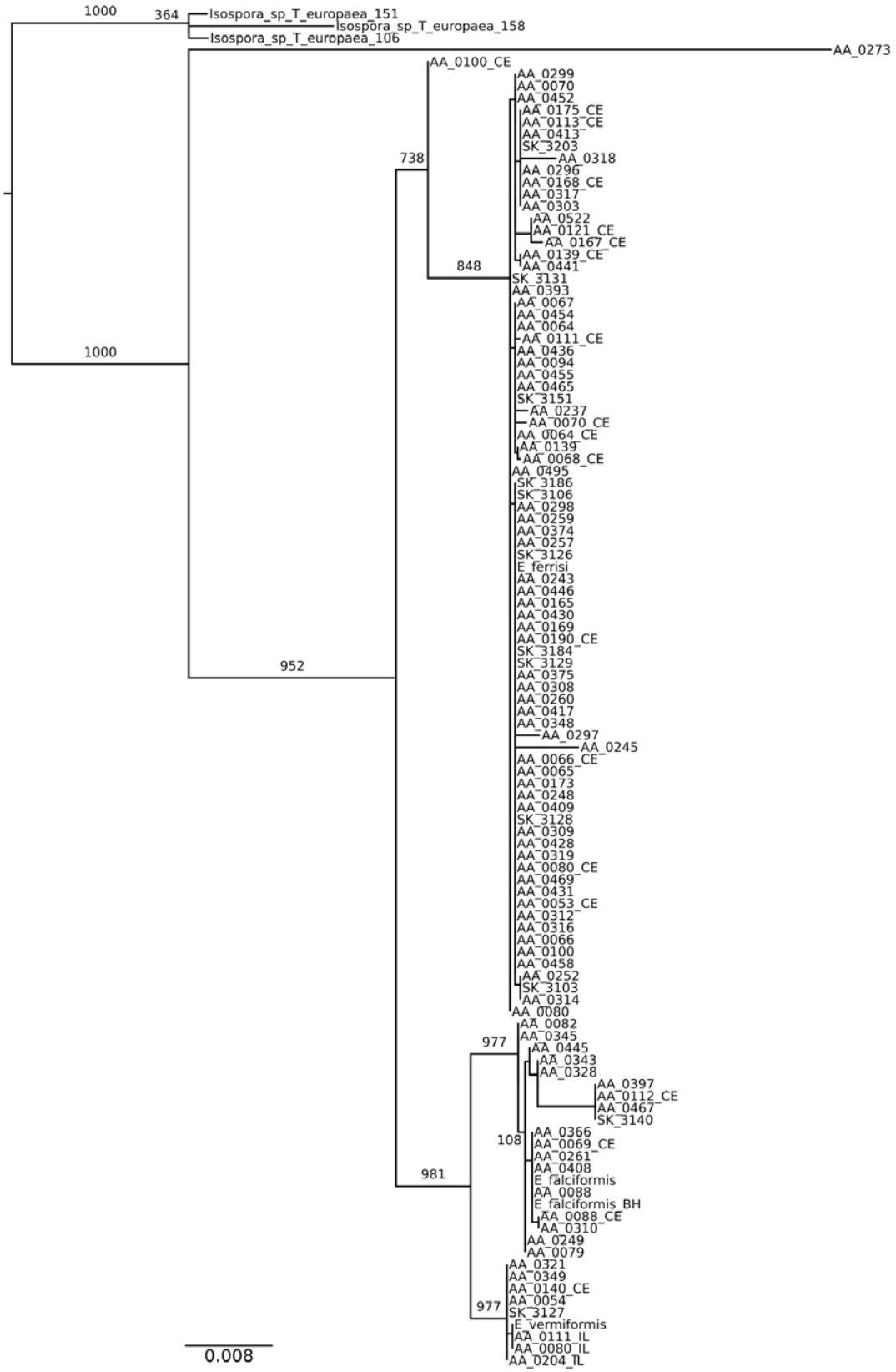
**Supplementary data S2.2 Accession numbers for 18S and COI sequences from *Eimeria* spp. wild isolates.**

**Supplementary data S2.5 Oocysts measurements from wild-derive isolates of *Eimeria falciformis, E. ferrisi and E. vermiformis***

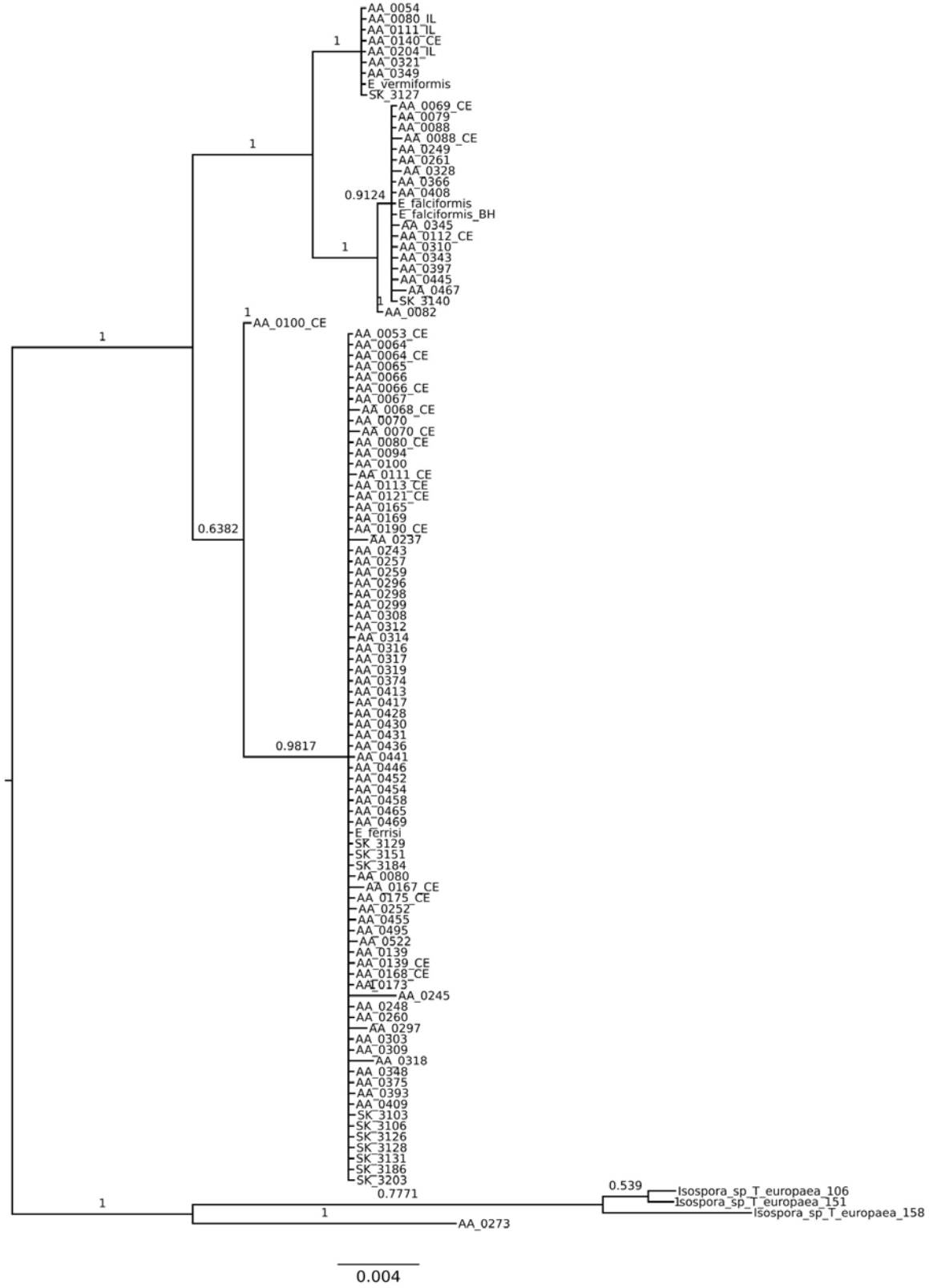
**Supplementary data S2.6 Ct values from the qPCR detection of *Eimeria* in tissue.**

All these information is available and can be downloaded at:
https://ars.els-cdn.com/content/image/1-s2.0-S2213224419301105-mmc1.xlsx

**Supplementary data S2.3 Maximum-Likelihood phylogenetic tree inferred for 18S/COI sequences supermatrix.**



**Supplementary data S2.4 Bayesian phylogenetic tree inferred for 18S/COI sequences supermatrix.**



The following supplementary materials from **Chapter 3** correspond to tables containing raw data not included in the present thesis:

**Supplementary data S3.1 Origin of samples.** Available at**:**

https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Fece3.5992&file=ece35992-sup-0001-Supinfo.xlsx

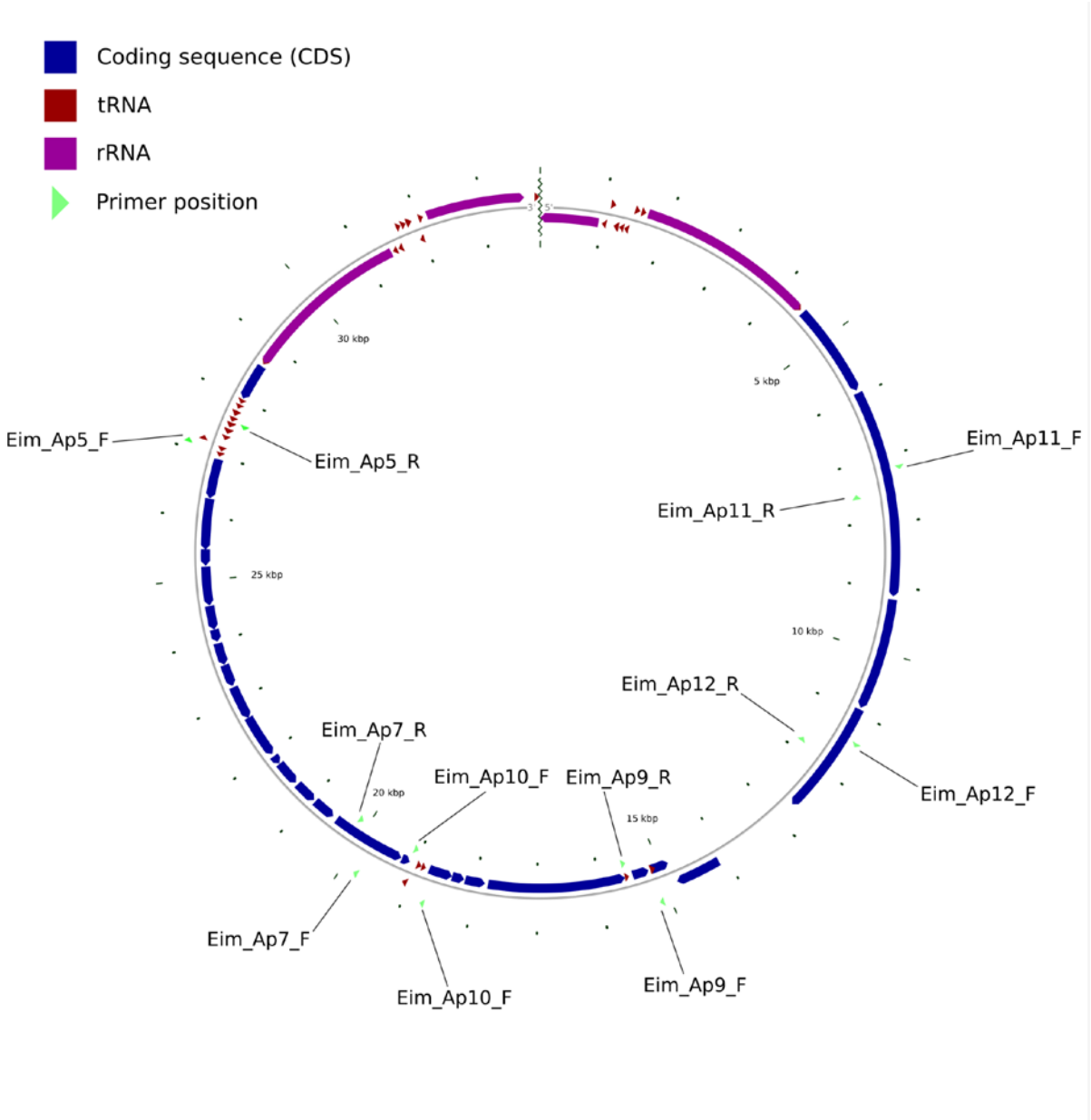
**Supplementary data S3.2 Accession number from reference sequences used for phylogenetic analysis.** Available at:

https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Fece3.5992&file=ece35992-sup-0002-Supinfo.xlsx

**Supplementary data S3.3 Primer pair list for multimarker genotyping of *Eimeria*.** Available at:

https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Fece3.5992&file=ece35992-sup-0003-Supinfo.xlsx


**Supplementary data S3.4 Position of primers for genotyping in *Eimeria falciformis* nuclear genome.** Primer pairs designed to amplify protein coding exons from *E. falciformis* were located across the nuclear genome of *E. falciformis*. Primer pairs are marked with yellow marks and names are highlighted in blue on the right side of the contigs. When more than one primer pair is located in the same contig, names are ordered from left to right.

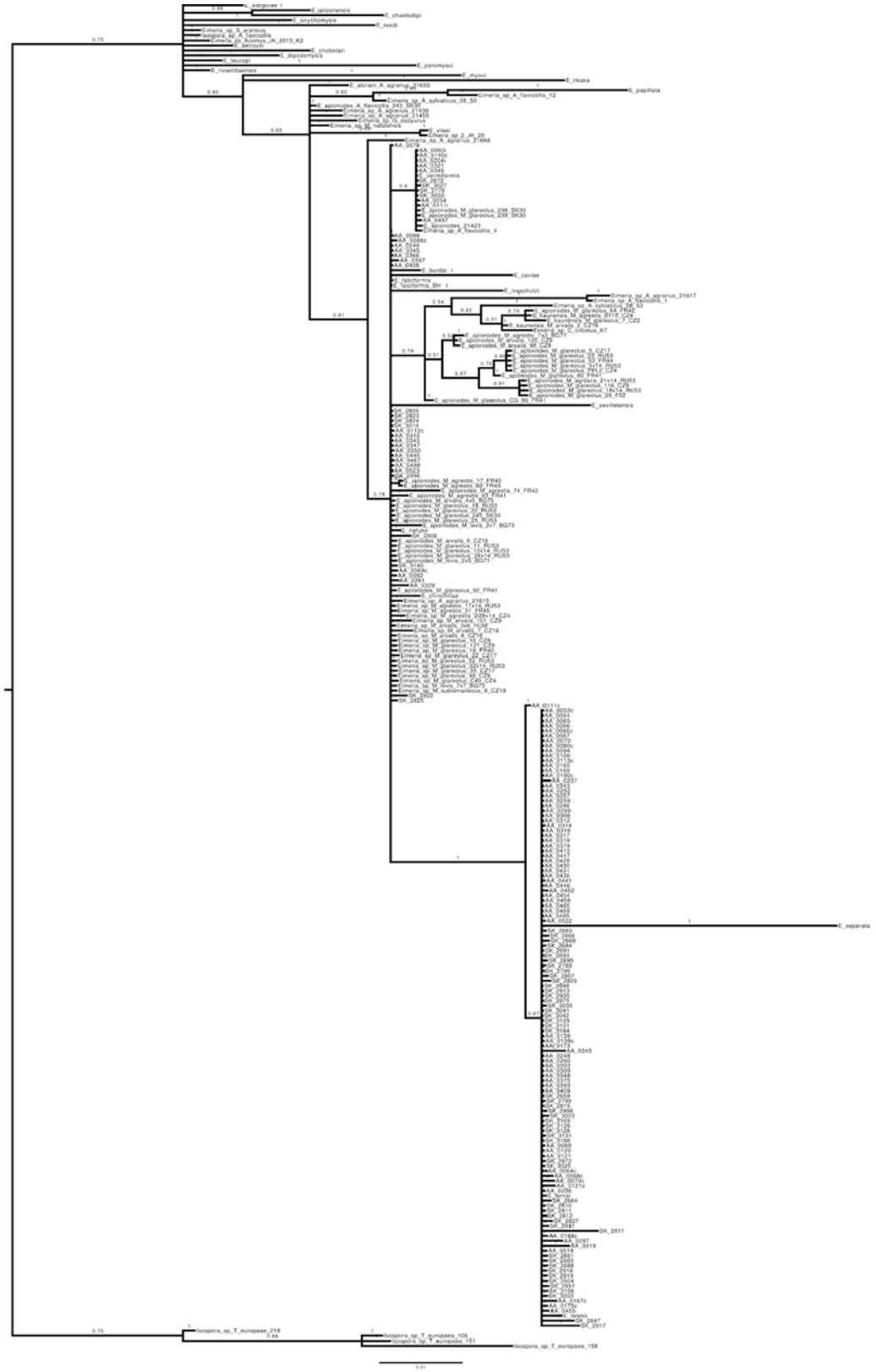



**Supplementary data S3.5 Position of primers for genotyping in *Eimeria falciformis* apicoplast genome.** Primer pairs designed to amplify different regions from the apicoplast of *Eimeria* spp. were located across the circular annotated representation of this genome. Primer pairs are marked with green marks and names are highlighted.

**Supplementary data S3.6 Amplified Sequence Variants (ASVs) recovered for each sample in multilocus amplification.** Each heatmap represents one marker, samples are listed in rows, different ASVs in columns. The number of sequence reads recovered per ASV and sample is log10 transformed and displayed in a colour shading. Pages 1-10: Apicoplast regions, 11 – 75: nuclear makers. Available at:

https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Fece3.5992&file=ece35992-sup-0006-Supinfo.pdf

**Supplementary data S3.7 Concatenated phylogenetic inference with nuclear (18S rDNA), mitochondrial (COI) and apicoplast (ORF470) sequences.**



**Supplementary data S3.8 Estimation of informative sites within the multilocus alignment of nuclear genes.** Available at:

https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Fece3.5992&file=ece35992-sup-0008-Supinfo.txt

**Supplementary data S3.9 Estimation of informative sites within the multilocus alignment of apicoplast regions.** Available at:

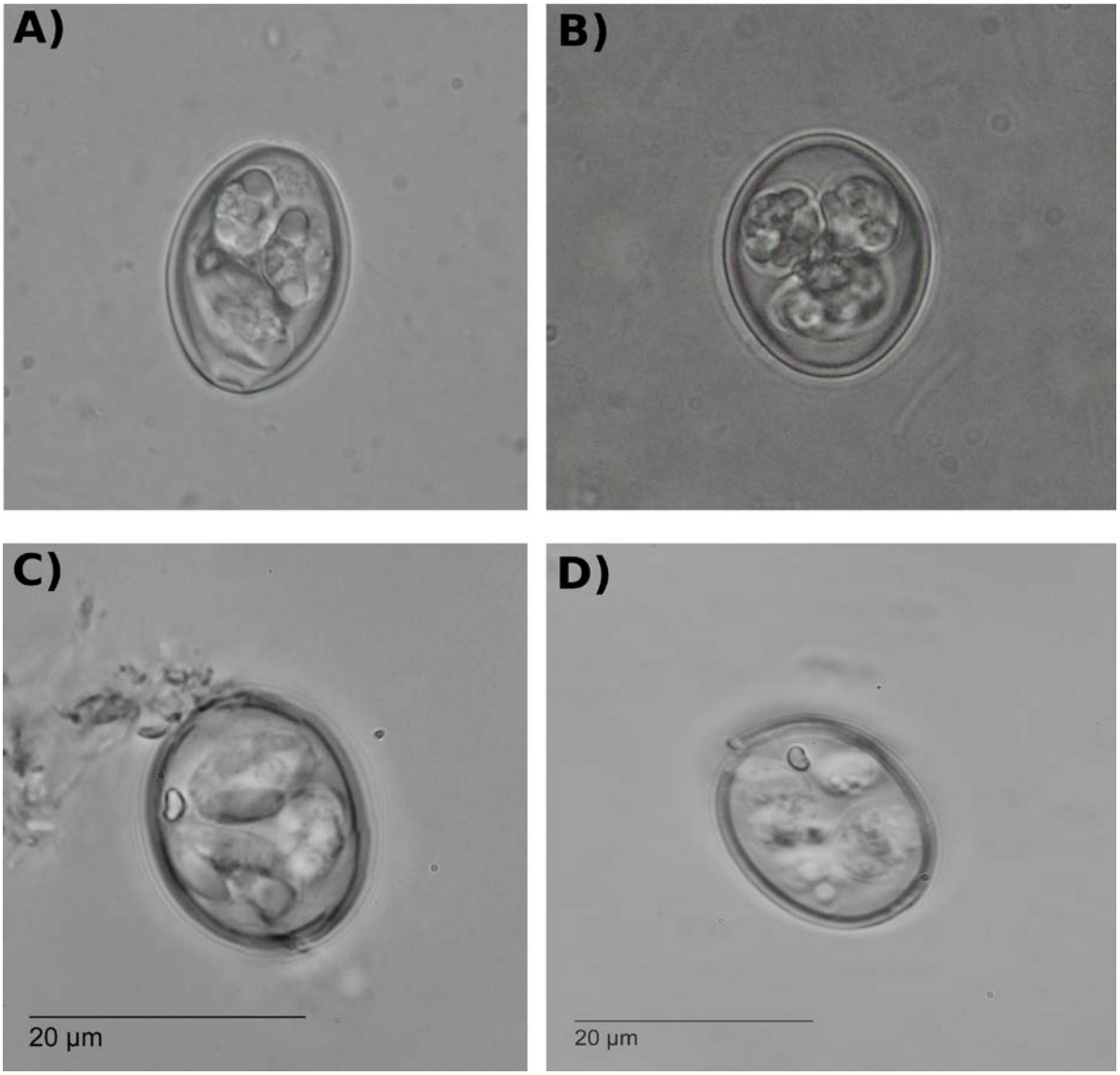
https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Fece3.5992&file=ece35992-sup-0009-Supinfo.txt

**Supplementary data S3.10 Morphology comparison between *Eimeria* species from *Mus musculus* and *Apodemus* spp.**

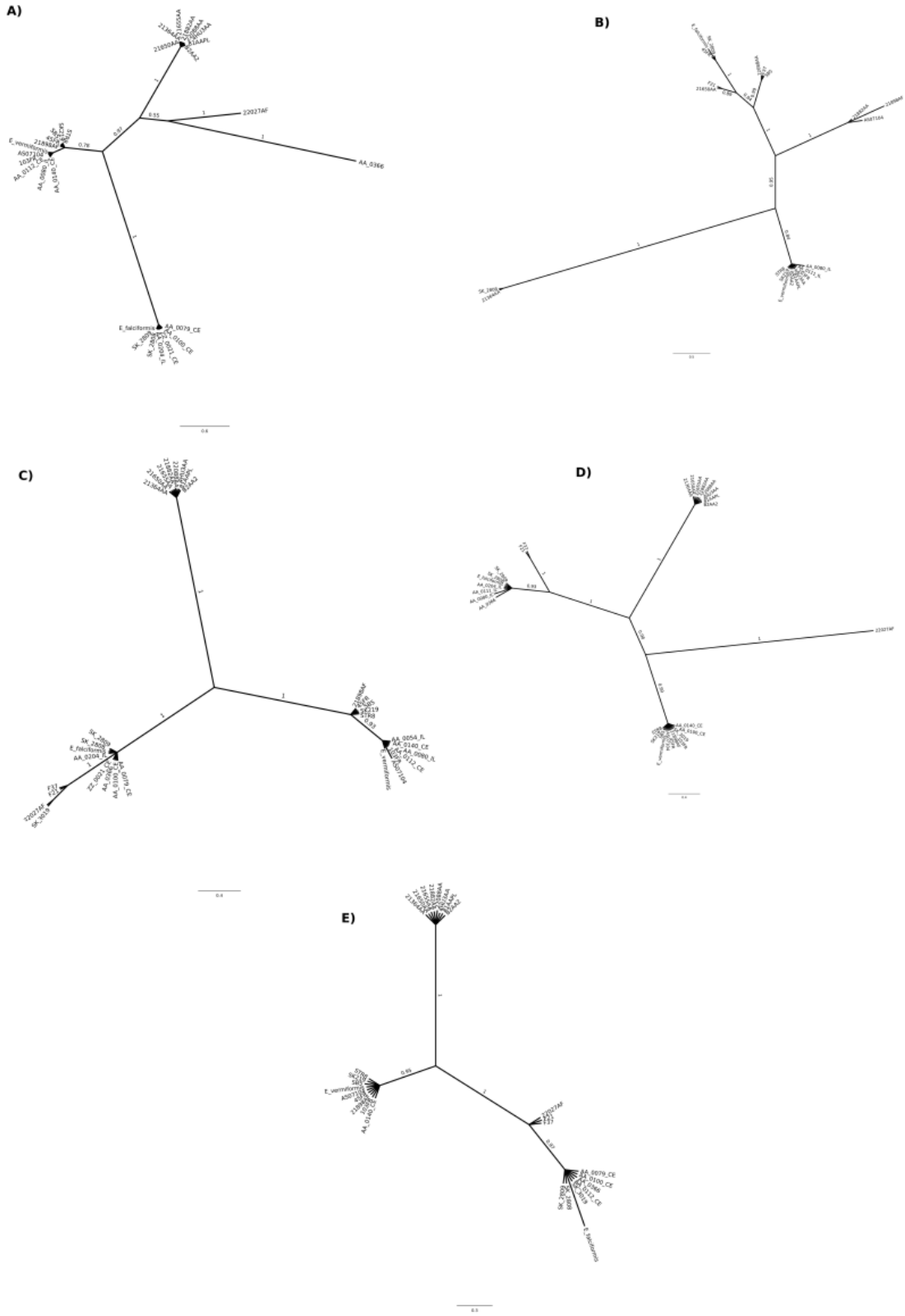
| Species of Eimeria | Host (s) | Shape of oocyst | Oocyst wall | Oocyst dimensions (µm)§ | Sporocyst dimensions (µm) | Oocyst residuum | Polar granule | Reference |
|---|---|---|---|---|---|---|---|---|
| *E. alorani** | *Aa* / A.m* | Ellipsoidal | Smooth | 26.9x19.3 (23-29)x(18-22) | 12.9x7.7 (12-14)x(7-9) | Absent | Present | Mácová et al. 2018; Hůrková et al. 2005 |
| *E. apionodes** | *Af* / Aa/ As* / Au / Mg* | Short piriform | Smooth | 20x17 (17-23)x(13-18) | 12x8 | Absent | Absent | Mácová et al. 2018; Pellérdy, 1974 |
| *E. falciformis** | *Mm* | Spherical / ellipsoidal | Smooth | 18.6x15.9 (15-21)x(13-18) | 8.8x5.5 (7-10) | Absent | Present | Jarquín-Díaz et al. 2019 |
| *E. falciformis* | *Mm* | Ovoid / spherical | Smooth | 21x18 (15-26)x(13-24) | 11x7 (10-12)x(6-8) | Absent/ Present | Present | Eimer, 1870; Haberkorn, 1970 |
| *E. vermiformis** | *Mm* | Spherical / ellipsoidal | Smooth | 20x15.6 (16-23)x(12-18) | 20x15.6 (16-23)x(12-18) | Absent | Present | Jarquín-Díaz et al. 2019 |
| *E. vermiformis* | *Mm* | Spherical / ellipsoidal | Lightly pitted | 23.1x18.4 (18-26)x(15-21) | 12.8x7.9 (11-14)x(6-10) | Absent | Present | Ernst et al. 1971 |

* Isolates used in the current study; § Numbers in brackets represent the range

Aa= *Apodemus agrarius;* Am= *A. mystancinus*; Af= *A. flavicollis*; Au= *A. uralensis*; As= *A. sylvaticus*; Mg= *Myodes glareolus*; Mm= *Mus musculus*




**Supplementary data S3.11 Micrographs of sporulated oocysts of *Eimeria* species from *Mus musculus* and *Apodemus* spp.** A) *Eimeria alorani,* B) *E. apionodes,* C) *E. falciformis* and D) *E. vermiformis* (Take from Jarquín-Díaz *et al.*, 2019). Scale bar=20 µm, all in the same scale.




**Supplementary data S3.12 Individual phylogenetic trees inferred from apicoplast markers.** Trees inferred for the apicoplast genetic markes in an analysis of individual regions. A) Ap7, B) Ap10, C) Ap11, D) Ap12 and E) Ap 5.

# List of publications and authors contributions

This dissertation is based on the following manuscripts:

**Chapter 2**

---

**Jarquín-Díaz VH**, Balard A, Jost J, Kraft J, Dikmen MN, Kvičerová J, Heitlinger E (2019). Detection and quantification of house mouse *Eimeria* at the species level – challenges and solutions for the assessment of Coccidia in wildlife. *IJP: Parasites and Wildlife*. doi: 10.1101/636662.

**Author contributions:**

VHJD and EH designed the project and obtained funding. VHJD, AB, JJ, JK and MND obtained data, VHJD led the data analysis and interpreted the results with support of EH, AB and JK. VHJD and EH wrote the manuscript with contributions from all other authors. EH supervised the project.

**Chapter 3**

---

**Jarquín-Díaz VH**, Balard A, Mácova A, Jost J, Roth von Szepesbéla T, Berktold K, Tank S, Kvičerová J, Heitlinger E (2020). Generalist *Eimeria* species in rodents: Multilocus analyses indicate inadequate resolution of established markers. *Ecol Evol* 00:1–12. doi: 10.1002/ece3.5992

**Author contributions:**
VHJD and EH designed the project and obtained funding. VHJD, AM, TRS, JJ, KB, ST and JK obtained data, VHJD, AB and EH designed the analysis, VHJD, AB and EH performed the analysis. VHJD, EH and JK interpreted the results. VHJD and EH wrote the manuscript with contributions from all other authors. EH supervised the project.

Other publications not related to this thesis are included in the CV.

# Declaration (Eidesstattliche Erklärung)

Ich erkläre hiermit, dass ich die vorliegende Dissertation selbständig abgeschlossen habe. Ich habe keine anderen als die angegebenen Hilfsmittel verwendet.
Ich erkläre hiermit, dass ich an keiner anderen Universität ein Prüfungsverfahren beantragt habe und dass ich das vorliegende Werk nicht als Dissertation bei einer anderen Fakultät eingereicht habe. Alle in Zusammenarbeit erarbeiteten und erzielten Ergebnisse sind entsprechend gekennzeichnet.
Die vorliegende Dissertation wurde in der Gruppe Ökologie und Evolution molekularer Parasit-Wirt-Interaktionen am Leibniz-Institut für Zoo- und Wildtierforschung (IZW), Berlin und an der Humboldt-Universität zu Berlin (HU) an der Fakultät für Lebenswissenschaften am Institut für Biologie in der Abteilung Molekulare Parasitologie unter der Leitung von Prof. Dr. Emanuel Heitlinger erstellt.

I hereby declare that I have independently completed the present dissertation. I have not used other aids than those indicated.
I hereby declare that I have not applied for an examination procedure at any other university or submitted to another faculty as a dissertation. All results produced and achieved in cooperation are indicated accordingly.
The present dissertation was produced in the group of Ecology and Evolution of molecular Parasite-Host Interactions at the Leibniz Institute for Zoo- and Wildlife Research (IZW), Berlin and at Humboldt University of Berlin (HU) at the Faculty of Life Sciences at the Institute for Biology in the Department of Molecular Parasitology, under the supervision of Prof. Dr. Emanuel Heitlinger.

Berlin, 25th September 2020 Víctor Hugo Jarquín-Díaz